The Application of Machine Learning to At-Risk Cultural Heritage Image Data

ROBERTS, MATTHEW,IAN

The Application of Machine Learning to At-Risk Cultural Heritage Image Data

Matthew Roberts

A Thesis presented for the degree of
Master by Research

Department of Archaeology
University of Durham
England
November 2019

The Application of Machine Learning to At-Risk Cultural Heritage Image Data

Matthew Roberts

Submitted for the degree of Master by Research
November 2019

Abstract

Abstract

This project investigates the application of Convolutional Neural Network (CNN) methods and technologies to problems related to At-Risk cultural heritage object recognition. The primary aim for this work is the use of developmental software combining the disciplines of computer vision and artefact studies, developing applications in the field of heritage protection specifically related to the illegal antiquities market. To accomplish this digital image data provided by the Durham University Oriental Museum was used in conjunction with several different implementations of pre-trained CNN software models, for the purposes of artefact Classification and Identification. Testing focused on data capture using a variety of digital recording devices, guided by the developmental needs of a heritage programme seeking to create software solutions to heritage threats in the Middle East and North Africa (MENA) region. Quantitative data results using information retrieval metrics is reported for all model and test sets, and has been used to evaluate the models predictive results.

Table of Contents

List of Figures

ACRONYMS/ABBREVIATIONS

CIDOC	International Committee for Documentation
CMS	Collection Management System
CNN	Convolutional Neural Networks
COTS	Custom-Off-The-Shelf
CRISP-DM	Cross-industry standard process for data mining
GPU	Graphics Processing Units
ICOM	International Council of Museums#
JPL	Jet Propulsion Laboratory
MEC	Model Export Certificate
MIT	Massachusetts Institute of Technology
ML	Machine Learning
SQL	Structured Query Language
UNESCO	United Nations Educational, Scientific and Cultural Organization (UNESCO)
UNIDROIT	International Institute for the Unification of Private Law
VASARI	Visual Arts System for Archiving and Retrieval of Images
WCO	World Customs Organization

DECLARATION OF AUTHORSHIP

I, Matthew Roberts, declare that this thesis titled, "The Application of Machine Learning to At-Risk Cultural Heritage Image Data" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: ____

Date: ____

Acknowledgements/Dedication

This thesis would not have been possible without the resources from the Durham University Oriental Museum, and I would like to give my most sincere gratitude to Helen Armstrong, Rachel Barclay, Craig Barclay, and Charlotte Spink for all their support these past several years.

Thanks to Dr. Kamal Badreshany and Dr. Marco Nebbia for their guidance in beginning this work, and to my supervisor Professor Anna Leone for the knowledge and encouragement necessary to see it to its end. Thanks to the Training in Action and the Heritage Documentation and Protection projects for support material and equipment.

Thanks to Dr. Boguslaw Obara, Dr. Chris Willcocks, Dr. Noura Al Moubayed, Stephen Bonner, and John Brennan from the Computer Science department for all their technical reviews and insights. Thanks to the CS postgraduate researcher crew for advice, support, and late-night Karaoke.

Thanks to Ustinov College for providing friends and family far from home.

Thanks to Mom; not all who wander are lost.

Thanks to Dad; no matter where you go, there you are.

Glossary[1]

Activation Function - A function which takes in the weighted sum of all of the inputs from the previous layer and then generates and passes an output value to the next layer.

B**ackpropagation -** The primary algorithm for performing gradient descent for a neural network. First, the output values of each node are calculated in a forward pass. Then, the partial derivative of the error with respect to each parameter is calculated in a backward pass.

Batch - The set of examples used in one iteration (that is, one gradient update) of model training.

Bias - An intercept or offset from an origin. Bias (also known as the bias term) is referred to as *b* or w_0 in machine learning models.

Class - One of a set of enumerated target values for a label. For example, in a binary classification model that detects spam, the two classes are spam and not spam. In a multi-class classification model that identifies dog breeds, the classes would be poodle, beagle, pug, and so on.

Convergence - Informally, often refers to a state reached during training in which training loss and validation loss change very little or not at all with each iteration after a certain number of iterations. In other words, a model reaches convergence when additional training on the current data will not improve the model.

Epoch - A full training pass over the entire dataset such that each example has been seen once. Thus, an epoch represents N/batch size training iterations, where N is the total number of examples.

Feature - An input variable used in making predictions.

Gradient Descent - A technique to minimize loss by computing the gradients of loss (a vector of partial derivatives of the model function, pointing in the direction of steepest ascent) with respect to the model's parameters, conditioned on training data. Informally, gradient descent iteratively adjusts parameters, gradually finding the best combination of weights and biases to minimize loss.

Learning Rate - A scalar used to train a model via gradient descent. During each iteration, the gradient descent algorithm multiplies the learning rate by the gradient. The resulting product is called the gradient step.

Model - The representation of what a machine learning system has learned from the training data.

Neural Network - A model that, taking inspiration from the brain, is composed of layers (at least one of which is hidden) consisting of simple connected units or neurons followed by nonlinearities (i.e. calculations which do not vary outputs in proportion to their inputs).

Neuron - A node (connected unit) in a neural network, typically taking in multiple input values and generating one output value. The neuron calculates the output value by applying an activation function to a weighted sum of input values.

[1] Definitions taken from the Google Developers Machine Learning Glossary https://developers.google.com/machine-learning/glossary

Normalization - The process of converting an actual range of values into a standard range of values, typically -1 to +1 or 0 to 1. For example, suppose the natural range of a certain feature is 800 to 6,000. Through subtraction and division, you can normalize those values into the range -1 to +1.

Overfitting - Creating a model that matches the training data so closely that the model fails to make correct predictions on new data.

Partial Derivative - A derivative in which all but one of the variables is considered a constant. For example, the partial derivative of $f(x, y)$ with respect to x is the derivative of f considered as a function of x alone (that is, keeping y constant). The partial derivative of f with respect to x focuses only on how x is changing and ignores all other variables in the equation.

Test Set - The subset of the dataset that you use to test your model after the model has gone through initial vetting by the validation set.

Training - The process of determining the ideal parameters comprising a model.

Training Set - The subset of the dataset used to train a model.

Validation - A process used, as part of training, to evaluate the quality of a machine learning model using the validation set. Because the validation set is disjoint from the training set, validation helps ensure that the model's performance generalizes beyond the training set.

Validation Set - A subset of the dataset—disjoint from the training set—used in validation.

Weight - A coefficient for a feature in a linear model, or an edge in a deep network. The goal of training a linear model is to determine the ideal weight for each feature.

1 INTRODUCTION

In the field of archaeological preservation, digital image data has become increasingly common in heritage documentation. Preserving heritage assets relies on recording as much accurate information as possible, both in service to the values of heritage custodianship (i.e. that the act of capturing a recording by itself has value) as well as a way to safeguard against potential threats to that heritage (vandalism, war, theft, natural degradation and many other potential destructive hazards). Accomplishing this is heavily reliant upon the tools available at the time of the heritage effort; nowadays, those toolsets might include digital cameras and tablets, remote sensing drones and many other methods of "born-digital" recordings. These tools in turn generate new types of documentation information, such as images, 3D models, and multispectral recordings, along with all the associated metadata. Their adoption has the potential to improve the work of heritage professionals by increasing the volume and ease with which asset data can be collected. However, it is this very potential which can also serve as the greatest challenge to heritage professionals, as the expertise necessary to properly utilize the tools and manage the information they produce can in turn overwhelm the resources of the project if not planned for appropriately. Vast amounts of data are only useful if a method exists to elicit meaningful information from that collection. The goal of automating the search of digital images in particular has been a cross-disciplinary research goal for many diverse fields, all with the ultimate intent of achieving better methods of data retrieval; in Computer Science and Data Engineering, this has traditionally fallen under the category of Computer Vision and Image Processing (da Silva Torres and Falcao 2006) (Liu, et al. 2007).

1.1 COMPUTER VISION AND IMAGE PROCESSING

The beginnings of meaningful digital image processing and computer vision tasks took place in the 1960's and 70's, driven in large part by the development of advanced computer systems by both the space program (such as the Jet Propulsion Lab Ranger-7 transmitting images of the moon in 1964) as well as new techniques and technologies in the medical imaging profession (the development of computerized axial tomography for medical diagnosis, for example) (Gonzalez and Woods 2006, 3-7). Spurred forward by this progression of technology, academic research began to explore the potential applications and capabilities of computer vision. These initial efforts were in many cases more ambitious rather than achievable, given the technology of the time; a prime example of this is the Massachusetts Institute of Technology (MIT) Summer Vision project (Papert 1966), whose stated purpose was "an attempt to use our summer workers effectively in the construction of a significant part of a visual system" and whose final goal was "OBJECT IDENTIFICATION which will actually name objects by matching them with a vocabulary of known objects." Even at this early stage of development researchers were considering the possibilities for computer enhanced photography to support artefact analysis (Burton, et al. 1970). Archaeologists and mathematicians with an interest in the use of computer applications began to pursue common cause and organization, with the first Computer Applications in Archaeology conference being held in Birmingham in 1973 (J. D. Wilcock 1973). Broadly speaking, much of the published work from this period focused on using computer imaging in relation to aerial photography, artefact shape imaging (usually for pottery profiling), and the replacement of traditional hand-drawn site and artefact mapping with software drafting tools (Ryan 1988, 16-18). Moreover, this type of work was constrained by researchers with access to mainframe computer facilities, and as such was primarily limited to western institutions. It wasn't until the 1980's that archaeologists began to methodically explore the possibilities of computer graphics and their potential application to archaeological science, in large part due to the availability of microcomputers to individual archaeologists (J. D. Wilcock 1999). However, the technology necessary for the implementation and support of digital archaeology was still not widespread in terms of availability and scope. The catalyst necessary to see wider adoption only came about when digital cameras and scanners became commercial commodities.

1.1.1 Digitization and Reconstruction of Heritage

The early 1990s marked the point at which digital image capture began to replace traditional photographing techniques. In the wider business environment, steady effort was being made to convert physical paper records to digital as part of cost saving measures; digital imaging of the time was focused on how to then create systems capable of providing recognition and retrieval of these document’s textual and image components (Cullen, Hull and Hart 1997). Given this business-oriented focus, it is unsurprising that the first regular usage of digital imaging related to heritage came about from the commercial art world, rather than in museums and the field (Hamber 2006). At the beginning of the decade digital camera technology was still in its infancy and had not yet become a commonly held item of personal use. The Auction House Christie's was one of the first institutions concerned with photography of heritage objects to devote research and development towards implementing a fully digital replacement of traditional photography. Post-processing was the main developments here, balancing out image capture against the choice of color profiles. At the time commercial and industrial standards were only beginning to be discussed, most notably the specification produced by the International Color Consortium (a group formed by Adobe, Agfa, Apple, Kodak, Microsoft, Silicon Graphics, Sun Microsystems, and Taligent in 1993) (International Color Consortium 1996). While there were numerous challenges faced by the implementation of the new technology, the project nevertheless showed that the new born-digital workflow sped up process flow, improved quality, reduced costs, and created the opportunity for new services.

1.1.1.1 Antique Books and Manuscripts

The use of computer vision to help with reading faded or damaged historical manuscripts is one of the earliest examples of digital imaging being used in heritage work; the research performed by historians at the California Technical Institute (Dr. John Benton, J. M. Soha, and A. R. Gillespie) are an early demonstration of this development (Benton, Gillespie and Soha 1979). Utilizing image enhancing techniques developed at JPL to restore and clean up images received from Mars (including contrast enhancement and spatial filtering), they managed to restore previously illegible medieval documents and advance the state of discussion for sections previously in dispute. However, after a decade of evaluation, the team's conclusion in the 1980's was that digital imaging was not yet commercially viable for manuscript studies (Kiernan 1991). Nevertheless, other heritage experts inspired by their work continued to explore the usage of digital imaging, testing it against known problems in manuscript visualization. Again, this early work was in many ways limited by expensive equipment found in specialized research facilities (Caltech and the British Library, for example). Dedicated digital cameras were being commercially produced (primarily for medical imaging), and financing initiatives by top archive institutions were beginning to see the value in supporting digitization efforts (such as the British Library's Initiative for Access programme, producing the Electronic Beowulf (Prescott 1997)). The landmark effort to digitize the damaged and fragile Beowulf manuscript demonstrated several advantages that would become the focus of heritage efforts for the next decade; primarily this included increasing accessibility for research (combining efforts from the British Museum, Harvard, and the University of Kansas), improving workflows (UV and infrared lighting of the document which had been time and cost prohibitive previously suddenly became technically practical for all elements of the folio), and providing the raw digital material for image processing (restoring known missing letters lost to deterioration or hidden by previous conservation efforts, as well as removing discolorations and improving legibility through thresholding techniques). The project was, in fact, the foundational example used by the British Library to describe their new strategic direction to support digital collaborations in research and scholarship (Brindley 2002).

1.1.1.2 Easel Paintings and Photographs

Given the enthusiasm for the potential benefits of digital imaging, it is unsurprising that parallel efforts were also being developed for paintings, photographs, and other similar two-dimensional heritage objects. One of the earliest initiatives to replace traditional photography with a digital system was the European visual arts system for archiving and retrieval of images (VASARI) project (Martinez, Cupitt, et al. 2002), initially funded between 1989 and 1992. Installed both in the National Gallery in London as well as the Doerner Institute in Munich, its purpose was to demonstrate the feasibility of high-resolution digital image processing techniques (Martinez 1991). As with the Initiative for Access programme, many of the key advantages of digital restoration were demonstrated by this project: experiments with a virtual copy do not damage the original, can be more easily supplied to researchers, and have the potential of automation for algorithmic corrections (Stanco, Restrepo and Ramponi 2011). Moreover, these types of objects suffer similar degradation issues as manuscripts, including aging discolouration (chemical, fungal, metal-induced, etc) or staining. Over the course of a decade, the project was used to aid several areas of conservation research, including detection of colour change over a period of time, damage resulting from transportation, simulation of conservation efforts, and superposition of multispectral images for comparative studies. Further research extended upon those initial ideas. Colour pigments that had faded could be returned to their former appearance through pixel hue transformations (Dik, et al. 2002), while crack development (craquelure) from improper storage/transport or internal stress could be automatically detected and removed (Giakoumis and Pitas 1998).

1.1.2 Computer Vision and Museum Collections

The move towards digital brought with it an explosion in the field of image processing techniques, as researchers sought new ways to organize and classify these new archives of digital content (Smeulders, et al. 2000). Colour, shape, and texture all became an area of focus, and effort was made to create methods of object recognition in digital images using local feature description and detection algorithms. Computer vision sought new approaches to region and patch analysis, focused on identifying key points for object recognition which were invariant to changes in scale, viewpoint, and obfuscation (due to clutter/noise). When detecting point or feature (one which persisted over multiple scales), the patch or neighbourhood surrounding it could then be extracted and composed into a feature vector. This in turn could be used for various methods of comparison, often as a histogram of gradients or features. These types of traditional feature engineering algorithms were in turn adapted to the heritage sector, most notably as visual guides for museums. Research which followed these initial efforts has primarily followed the same pattern of implementation, using some variety of feature extraction and engineering software based on interest point detection (Föckler, et al. 2005) (Bay, Fasel and Gool 2006) (Ruf, Kokiopoulou and Detyniecki 2008) (Blighe, et al. 2008). Others have sought to use further specialized methods of image processing to classify according to art type or cultural provenance, including combinations of image segmentation techniques (Haladova 2010) and statistical analysis of edge orientations (Redies, Brachmann and Wagemans 2017). Nevertheless, half a century after MIT's announcement that they would solve Object Identification over the course of the summer, that task remains an aspirational goal within the computer vision field. The fundamental difficulty continues to be the inability for algorithmic processing of raw digital image data to accurately achieve the sort of high level (or semantic) descriptions which a human observer would comprehend and label a visual object (Hare, et al. 2006). Digital image data is stored as a numerical representation, and even slight variations in the recording of an object can change those representational numbers entirely. Such challenges in the domain of image recognition include viewpoint variation, illumination, deformation, occlusion, and background clutter. Traditional image processing initially sought to address this by developing algorithms and image-processing techniques of increasing complexity, with researchers competing with one another in global challenges intended to refine and prove the effectiveness of a particular implementation (Valentino-DeVries 2011). From those competitions a specific methodology has grown increasingly popular over the last ten years: Deep Learning.

1.2 Convolutional Neural Networks (Deep Learning)

In 2012 an image recognition system designed by Alex Khrizevsky demonstrated a significant advancement in the ability of software to classify large image datasets automatically (Krizhevsky, Sutskever and Hinton 2012). The record-breaking performance in accuracy was driven in part by improvements in Graphics Processing Units (GPU), making feasible the training of a deep convolutional (i.e. multilayered) neural network. Inspired by biological design, these deep architectures have been an area of research going as far back as the 1980's (LeCun, Bottou, et al. 1998), with additional popularization resulting from research into unsupervised learning in the late 2000's (Hinton, Osindero and Teh 2006) (Bengio, et al. 2007). However, it was Krizhevsky's demonstration of drastic reduction in error rates for feature recognition through Convolutional Neural Networks (CNN) combined with large quantities of labelled digital data and GPU processing power which motivated a rapid and wide-scale adoption of their use (LeCun, Bengio and Hinton 2015). The computer vision community has since supported a number of efforts to find new ways in which this particular software solution can be applied to other collections of image data. In the realm of heritage, there have been some initial attempts to solve the problem of recognition using these kinds of machine learning techniques. Classifying classical art paintings according to artistic style, artist, and time period is one such example, using the flat 2D images of digital painting representations (Mensink and Gemert 2014) (Bar, Levy and Wolf 2014) (Tan, et al. 2016). Similarly, recent work in numismatics has experimented with using CNNs to detect facial profiles and characteristic landmarks on ancient coinage, trained on images from both museums and online coin dealers (Kim and Pavlovic 2016) (Schlag and Arandjelovic 2017). These efforts demonstrate some of the possible ways to collect and curate training data from a heritage institution seeking to investigate new methods of automated identification. However, in terms of museum collections there still exist several other types of objects and antiquities which represent an as-yet untapped potential as a research focus and resource, with associated heritage protection issues which are in need of improved capabilities in automated vision tasks. Investigating that potential is the aim of this work, as described below.

1.3 Research Project Definition and Aims

This project investigates the application of current Convolutional Neural Network and Machine Learning (ML) technologies for the purposes of problems related to At-Risk cultural heritage object recognition. The primary research aims described in this report focus on a set of specific recognition tasks testing the use of CNN machine learning for two methods of automatic recognition, namely:

1. How well does a curated, pre-existing multi-perspective image data set of a museum collection work for artefact typology classification?
2. Can multi-image photography of individual museum objects be used to accurately identify a specific museum object?

This was done with a goal of investigating the practicality of a software solution combining the disciplines of computer vision and artefact studies in a way that would be useful for future applications in the field of heritage protection. To organize this work we utilized a common development methodology known as the Cross-industry standard process for data mining (CRISP-DM), an open-source process which describes the iterative stages used to develop a machine learning project (Shearer 2000). The specific implementation utilized is shown in Figure 1, and described in the following chapters:

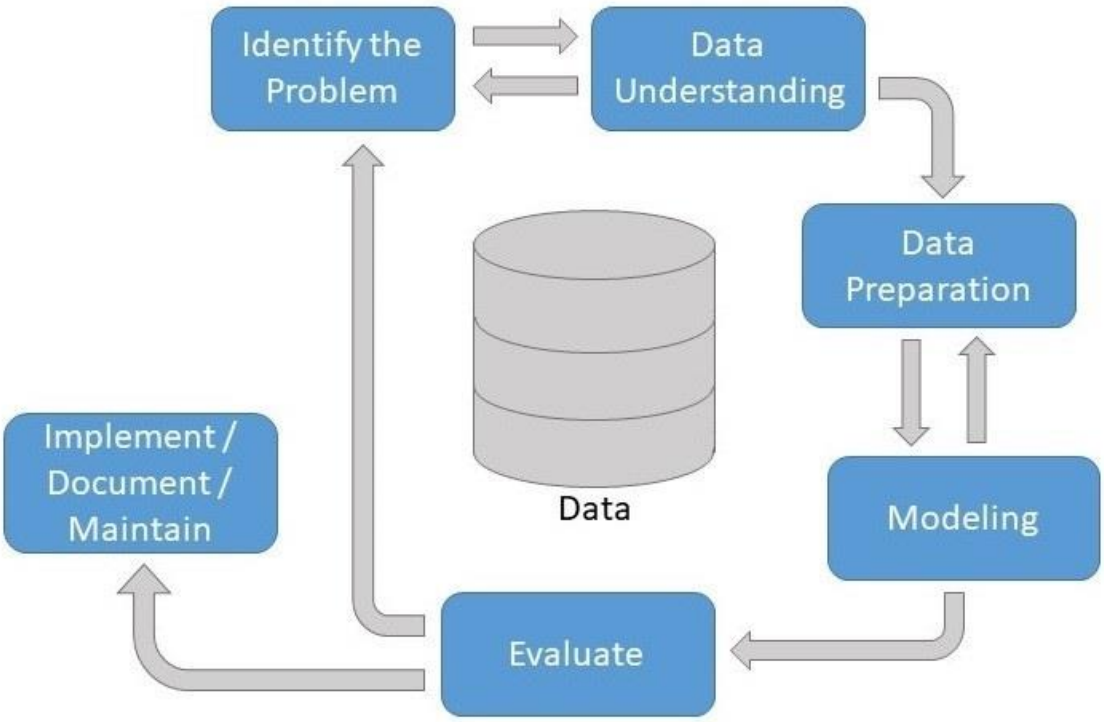

Figure 1 - CRISP-DM Diagram for Machine Learning Lifecycle, figure created based on original paper (Shearer 2000)

1. **Identify the Problem**: The first step of the life cycle is to define the problem to be solved using machine learning, along with the specific objectives. Threats to heritage stemming from the illegal antiquities market was our primary consideration when developing this work, the background and problem definition for which is discussed in **Chapter 2**.
2. **Data Understanding and Preparation**: The next step is to collect and prepare all the relevant data for the project. This means consulting domain experts to determine what data might be relevant, gathering that data, and organizing it into a format suitable for analysis. Digital image data provided by the Durham University Oriental Museum was collected from both a pre-existing catalogue of digital records (primarily consisting of their Egyptian and Asiatic collections), as well as by photographing previously unrecorded objects from their archives. The data acquisition and resulting data set creation is described in **Chapter 3**.

3. **Model Data and Evaluate:** After data acquisition the next phase is the development of software to process the data and create predictions according to the project goals, as well as defining the metrics by which the results shall be judged. Using the data from the Oriental Museum, various implementations of CNN software models for the purposes of artefact classification were developed, trained, and tested; quantitative data results using information retrieval metrics were collected for all models and test sets, and used to compare the model efficiencies. Our primary CNN model development is described in **Chapter 4**, focusing on Object Classification.
4. **Iterate**: Within this methodology, going back through the previous is considered key to building greater system capability. Returning to the original problem definition and development needs, a new data set was created from previously unrecorded Egyptian figurines originating from the Wellcome Collection. This data set was created using multi-image burst mode photography, and used to test the utility of the previously developed models for Object Identification. The training was further iterated upon with an attempt to combine the Identification and Classification data sets, incorporating training data for Egyptian Ushabti figurines drawn from the original image data set and tested using images retrieved from eBay and previous Oriental Museum ushabti photography work by students. This work is described in **Chapter 5**.
5. **Implement/Document/Maintain**: The final step is to implement, document, and maintain the data science project so the key stakeholders can continue to expand and improve upon it. The concluding chapter provides the formal assessment of our experimental results in terms of its potential use for a developed recognition system, as well as suggestions for how the results might inform future heritage digital imaging work projects.

2 Problem Definition - The Illicit Antiquities Market

2.1 International Agreements and Protection of Cultural Heritage

2.1.1 Ethical Frameworks for Collections

In the period of the late nineteenth and early twentieth century, archaeologists, collectors, and museum curators often sought to develop their collections through acquisitions made upon the open market, either through purchases at auction houses or directly from the regional source (Brodie 2012). According to the professional ethics of the time it was the material itself which was the most important factor for consideration, rather than the manner in which it was acquired. There was, however, a growing recognition internationally that heritage needed to be protected. In the wake of World War I and II the **1954 Convention on the Protection of Cultural Property in the Event of Armed Conflict** (ICOM Department of Programmes 1954) was adopted at The Hague, outlining a rough agreement regarding how heritage should be safeguarded during wartime. The meeting in which it was developed was considered significant, as it included both western and communist bloc countries. However, the final agreed upon declaration failed to get the support of the Anglo-Alliance nations (notably not being signed by either the US or the UK until 1965). Nevertheless it was an important first step in terms of defining the concept of cultural property, as well as outlining agreements necessary to safeguard said property through a lens of international relations (Meyer 1993). That said, “safeguarding” still included acquiring material and placing it in a (usually Western) museum; it did not matter if the acquisition was against the law in the country of origin, just so long as it met the legal obligations of the destination country. Brodie, Doole, and Watson summed up this ethical stance in their 2000 report for the International Council of Museums (ICOM), “If it was legal it was ethical, with some museums even arguing that open market acquisitions were necessary for those institutions without active programs of fieldwork and on-site research” (Brodie, Doole and Watson 2000).

As the century progressed, however, this ethical framework began to change, to reflect concerns beyond just the physical preservation of antiquities. A growing dissatisfaction had begun to take shape in the archaeological community as researchers found themselves increasingly frustrated by what was not known about items in their collections. Lack of details regarding an object's source were leaving holes in the archaeological record, inhibiting productive research. Moreover, the damage to archaeological sites from those supplying the antiquities trade was becoming an increasing concern; archaeologists began to question the role being played by western institutions in supporting a market that was resulting in ongoing destruction (Gerstenblith 2013). Governments of countries acting as the source of material began to have a greater awareness of their cultural heritage laws being violated, increasingly seeing the illegal excavation and export as direct attacks on state history and sovereignty. Museum's began to recognize that the undocumented material in their collections were physical gaps in the historical record, a loss of information about humanity's past which could never be reclaimed. A watershed moment came with the publication of Dr. Clemency Coggins article regarding the illicit trafficking of pre-Colombian antiquities (Coggins 1969), in which she documented not only the destructive removal of monumental Mayan stelae but also traced their destination into the possession of major US art museums (as well as private dealers and collectors). Coinciding with several journalistic exposes and museum controversies over smuggled art (Bator 1982), this finally created sufficient international pressure to force a response.

2.1.2 Response to Heritage Destruction as a Result of the Art Trade

The **Philadelphia Declaration** (Penn Museum 1980) was passed in 1970, wherein the University of Philadelphia's University Museum declared that it would no longer purchase unprovenanced art objects or antiquities specifically as a reaction against the looting and plundering of ancient sites (Pezzati 2010). Other museums began to follow suit, and shortly thereafter the first United Nations Educational, Scientific and Cultural Organization (UNESCO) conventions were drafted which specifically sought to establish ethical principles for the protection of cultural antiquities. The first (and most famous) of these was the **Convention on the Means of Prohibiting and Preventing the Illicit Import, Export and Transfer of Ownership of Cultural Property** (or in shorthand, the **1970 Convention**) (UNESCO 1970). The 1970 UNESCO Convention was one of the first international efforts aimed at arresting the growth of the trade in looted cultural antiquities. Both it and the Philadelphia Declaration are often cited as a watershed moment for professional ethics regarding the acquisition of cultural property. However, over the course of the next decade

UNESCO was confronted with the difficulty in implementation, specifically in adapting the convention to national laws (Prott 1996). To address this, UNESCO commissioned consultant organizations with expertise in private and commercial law to develop a supplementary convention handling the more difficult legal issues regarding illegally transported cultural objects. The result was the **1995 UNIDROIT Convention on Stolen or Illegally Exported Cultural Objects** (UNIDROIT 2014), produced by the UNIDROIT (International Institute for the Unification of Private Law) intergovernmental organization. The convention sought to establish uniform procedures for the restitution of stolen or illegally exported movable cultural objects, specifically those coming from "clandestine excavations." At its heart is the assertion that a possessor of a stolen artefact must return it whatever the circumstances. The UNIDROIT convention was intended to bring clarity to international law and the duties of signatories in addressing the market in illicit antiquities; however, the work was only considered a way to incrementally influence the art market to trend away from illicit antiquities over time, as the best workable outcome of international compromise. Accordingly, most of the subsequent analysis from those working in heritage protection were highly critical of both it and the 1970 Convention's ultimate effectiveness (Brodie and Doole 2001).

2.1.3 Response to Heritage Destruction as a Result of International Terrorism

More recently, the most significant impact on national agreements regarding the protection heritage was developed in response to destruction perpetrated under the aegis of international terrorism. In the months between June and October of 2014 a number of news articles were published whose headlines focused on the actions undertaken by the radical insurgent group Da'esh[2], many focusing on the group's role in selling looted antiquities as a means of funding their operations in the region. Several articles also sought to highlight the

[2] Also referred to as the Islamic State of Iraq and the Levant (ISIL), the Islamic State of Iraq and Syria (ISIS), and as the Islamic State (IS). This terrorist organization operated in the Iraq and Syria conflict zones, clashing with local governments, international armed forces, government opposition forces, and other insurgent groups. From 2014 to 2017 the group held and administered territory while claiming statehood, proclaiming itself as the start of a worldwide caliphate. In addition to their involvement in the illicit antiquities market, the United Nations also held them responsible for numerous human rights abuses, crimes against humanity, war crimes, and cultural/ethnic cleansing. Despite early successes in taking territory and establishing themselves as the de-facto power in several large cities, by the end of 2017 the proto-state had lost control of almost all of its regional holdings (and thus weakening its capability as a pipeline for looted antiquities). For a more in-depth analysis of the organization's historical role and actions during the 2014 period, see Zana Khasraw Gulmohamad's **The Rise and Fall of the Islamic State of Iraq and Al-Sham (Levant) ISI**S (Gulmohamad 2014) and Fawaz Gerge's **ISIS: A History** (Gerges 2017).

heritage community's efforts to save archaeological treasures in Syria. The list below shows a small sampling of high profile articles and headlines from major news outlets of the time, illustrating this sudden interest:

- *UNESCO Confirms ISIS Funding Terrorism by Selling Artifacts*, Ian Black, Rania Abouzeid, Mark Tran, Shiraz Maher, Roger Tooth and Martin Chulov, The Guardian, 16 June 2014
- *How Terrorists Tap a Black Market Fueled by Stolen Antiquities*, NBC News 23 June, 2014
- *ISIS Cashing in on Looted Antiquities to Fuel Iraq Insurgency*, Heather Pringle, National Geographic 27 June, 2014
- *The Quest to Save Syria's History*, Katrin Elger, Spiegel Online, 4 August, 2014
- *Illicit Trade in Looted Antiquities Helps Finance ISIS Terror Network*, Mark Vlasic, National Post, 15 September, 2014
- *UNESCO Confirms ISIS Funding Terrorism by Selling Artifacts*, Henri Neuendorf, Artnet news, 1 October, 2014
- *ISIS's Looting Campaign,* David Kohn, The New Yorker, 14 October, 2014
- *The Black-Market Battleground,* Justine Drennan, Foreign Policy 17 October, 2014

Soon afterwards, UNESCO issued two new security resolutions regarding the threat to heritage objects. The first was **Security Resolution 2199** (Resolution 2199 2015), adopted in February 2015. It details obligations of member states to take steps to prevent terrorist groups in Iraq and Syria from benefiting from trade in oil, antiquities and hostages, and from receiving donations. It also condemns any trade with Da'esh/ISIL, the Al-Nusrah Front, and any other entities associated with Al-Qaida. This was followed soon afterwards by **Security Resolution 2253** (Resolution 2253 2015), adopted in December 2015. This resolution focuses on topics such as freezing of assets, travel bans, arms embargoes, as well as listing criteria for ISIL, Al-Qaida and "associated individuals, groups, undertaking and entities" — measures that the Council decided it would review in 18 months or sooner, with a view to their possible strengthening. Finally, a further UNESCO resolution came in 2017, regarding heritage protection in a conflict zone: **Security Resolution 2347** (Resolution 2347 2017), adopted on 24 March 2017. It concentrates on the protection of cultural heritage in the event of armed conflicts, with the heritage looting witnessed in the countries of Syria, Iraq, and Mali specifically cited as the cause for this resolution. Calling back to earlier resolutions, it states that "Unanimously adopting resolution 2347 (2017), the 15-member Council recalled its condemnation of any engagement in trade involving Islamic State in Iraq and the Levant (ISIL/Da'esh), Al-Nusrah Front, and all other individuals or groups associated with Al-Qaida. It reiterated that such engagement could constitute financial support for entities designated by the 1267/1989/2253 ISIL (Da'esh) and Al-Qaida Sanctions Committee."

From these examples it is clear that the past decade has seen a renewed focus on the issues associated with the illegal antiquities trade, paired with a greater public awareness of heritage destruction. UNESCO and national governments continue to pass resolutions and seek international agreement regarding the protection of cultural heritage. Broad agreements are agreed to in response to global changes in the perception of what it means to conserve and protect cultural heritage. When it comes to evaluating the effectiveness of these resolutions ultimately, however, we are confronted by a problem which has persistently challenged those working in the field of heritage protection: lack of knowledge.

2.1.4 The Absence of Rigorous Data

In 1999 a symposium was held at Cambridge, with invitations sent out to archaeologists from around the world. The subject of the meeting was to discuss destruction of the world's archaeological heritage, and to specifically address the lack of clarity regarding the illicit antiquities market (Brodie and Doole 2001). According to the organizers, one of the primary difficulties in arguing for governmental response and action towards the market was the absence of a scientifically designed or systematic survey regarding the volume of the antiquities moving through the trade. Lacking such data, heritage professionals could not accurately categorize the damage it causes. Instead, arguments were reduced to anecdotal assertions based on personal experiences, often met by resistance from collectors and dealers profiting from the trade. The symposium's hope was that by inviting archaeologists from areas subject to the most drastic looting it would be possible to finally quantify the scope of the trade, to define the methods and relationships of illegal activities, and to move the discussion forward from anecdote to verifiable fact. Unfortunately, such was not the result:

"… from the papers presented at the Symposium, and now in this volume, it is clear that the information which it was hoped would emerge is often simply not available. There do not appear to be central, easily accessible registers of damage caused to archaeological sites. Even in relatively wealthy countries such as the United Kingdom and United States … creating a record of archaeological destruction is not regarded as a high priority, sometimes even within the archaeological community." (Brodie and Doole 2001)

A report commissioned by the International Council of Museums (ICOM) UK working in conjunction with the Museums Association and published at the same time went further, asserting that the vaguely defined nature of the trade is not simply a natural result from lack of interest but an intentional obfuscation from its participants (Brodie, Doole and Watson 2000). By keeping the details opaque, sellers can both maintain their market positions and argue that the illegitimate material on the market is either a small fraction of the overall trade or limited to minor unimportant items. The ultimate conclusion of the report was that addressing the looting of cultural material would require a top-down approach focusing on the collectors, museums, and dealers trading in unprovenanced material.

In the two decades since the situation has remained largely unchanged. In his chapter discussing trafficking of cultural objects within *The Oxford Handbook of Crime and Public Policy* (Tijhuis 2009), Antonius Tijhuis maintains that illegal transactions involving cultural artifacts remains the least studied form of transnational crime, with most of the interest coming from non-criminologists (archaeologists and journalists, primarily). He is critical of the multilateral treaties discussed above, believing that no evidence exists that their ratification has had any proven effect on the trade. He argues that the major limiting factor preventing an accurate judgement is that the size of the trade continues to remain unknown. This perspective is repeated in the more recent book on Heritage Crime (Polk 2014), in which Kenneth Polk re-iterates that though assertions exist stating that the trafficking of heritage material rivals other illicit markets in size, there is "literally no data that would support such assertions", and that "no body of crime statistics can be consulted to verify its size."

In contrast to this, there has been one notable source of new published data regarding trafficked cultural material. In 2012 the World Customs Organization (WCO) began to release an annual report aimed at providing in-depth analysis of global illicit trade through seizure data and case studies (World Customs Organization 2019). Data analyzed for the report is voluntarily submitted by members of the WCO[3]. Though the information in the reports from 2012 to 2017 regarding the illicit antiquities trade is based on limited seizure data which does not allow for a comprehensive analysis of global trend and patterns, it does provide an initial effort to quantitatively assess the smuggling of cultural material. Further, the information and structure of the reports year-by-year reflect a change in emphasis and attitude regarding the importance in assessing the illicit antiquities trade. In the 2012 report there was only a brief discussion of the illicit antiquities trade as a part of the Revenue section, limited to a single page heading titled *Trafficking of Cultural Goods* (World Customs Organization 2012, 33). Further, the 2013 and 2014 reports lacked even a single page for cultural heritage trafficking. This is contrasted by the 2015 report (notably following the period of attention stemming from international terrorism discussed in Section 2.1.3 above), in which an entire chapter is dedicated to the trade in illicit antiquities (World Customs Organization 2015, 136-146). Further, all subsequent reports continue this renewed focus, maintaining a full *Trafficking in Cultural Heritage* section comparable to other major areas of illicit materials trade. Reported data from customs control are visually shown, and a comparative analysis to the previous year's data discussed. This includes analysis on trends regarding types of items being trafficked, smuggling routes (from source to destination countries), methods being employed to move the illicit material, and a discussion of international group efforts to combat the trade (including the work of UNESCO, INTERPOL, and ICOM along with the WCO). Ultimately, the pattern which the WCO provides shows objects continuing to move across cultural borders which are suffering from regional instability, from conflict zone source countries towards more stable dealer markets. Moreover, they suggest that the trade is in a state of flux, with the last few years of data suggesting trafficking in smaller antiquities (specifically designated as coming from an archaeological site) most on the rise (World Customs Organization 2015) (World Customs Organization 2016) (World Customs Organization 2017). However, even this is caveated that

[3] Membership currently includes 183 countries, representing 98% of International Trade. (World Customs Organization 2019)

the trends emerge largely from seizure data submitted by countries in the Middle East, the Russian Federation, and Ukraine; while this could indicate that these places are intraregional and international hubs for the trade, it could also be attributed to the local governments using the trafficking of cultural heritage objects as a propaganda tool (Vikhrov 2018).

2.2 Organization of Supply Networks for Illicit Antiquities

"We do not feel at the present time that from an international perspective it is useful, or indeed possible, to distinguish between a 'licit' and 'illicit' trade in antiquities. Numerous case studies have shown that looted material is effectively 'laundered' as it passes through the trading network, so that what might be moved illegally out of one country will at a later date be offered for sale legally in a reputable outlet in another without that outlet knowing that the material was looted." (Brodie and Doole 2001)

In his chapter within the *Oxford Handbook of Public Archaeology*, Neil Brodie provides an overview of the methods by which heritage stakeholders have sought to understand and analyze the illicit antiquities market (Brodie 2012). Published academic work by archaeologists, anthropologists, and criminologists has primarily focused on the supply side of the trade, describing the methods and statistics of objects looted from source countries. By contrast, the dealers, collectors, and museums which are exposed as having acquired illicit material are generally detailed through the efforts of investigative journalism and media reports. Brodie suggests that this is in part due to the social power dynamics involved in the antiquities market: rich western collectors and museums are able to discourage academic inquiry utilizing threats of legal action. This is an intimidation tactic which can be more easily responded to by media companies, who are supported by legal departments better prepared to respond to legal posturing. Nevertheless a combination of provenance and market research, published media investigations, and ethnographic/criminologist surveys have provided a broad picture of the global structure of trafficking networks. In some analyses the trafficking is represented as a hierarchical set of layers; such a pyramid-like structure places looters at the bottom and collectors at the top, separated by various layers of middlemen (Massy 2008) (Kersel 2007) (Kaiser 1991). More recent researchers, however, have argued that this creates a false conception of the relationships for the participants involved. Instead, they maintain that it would be more accurate to view the

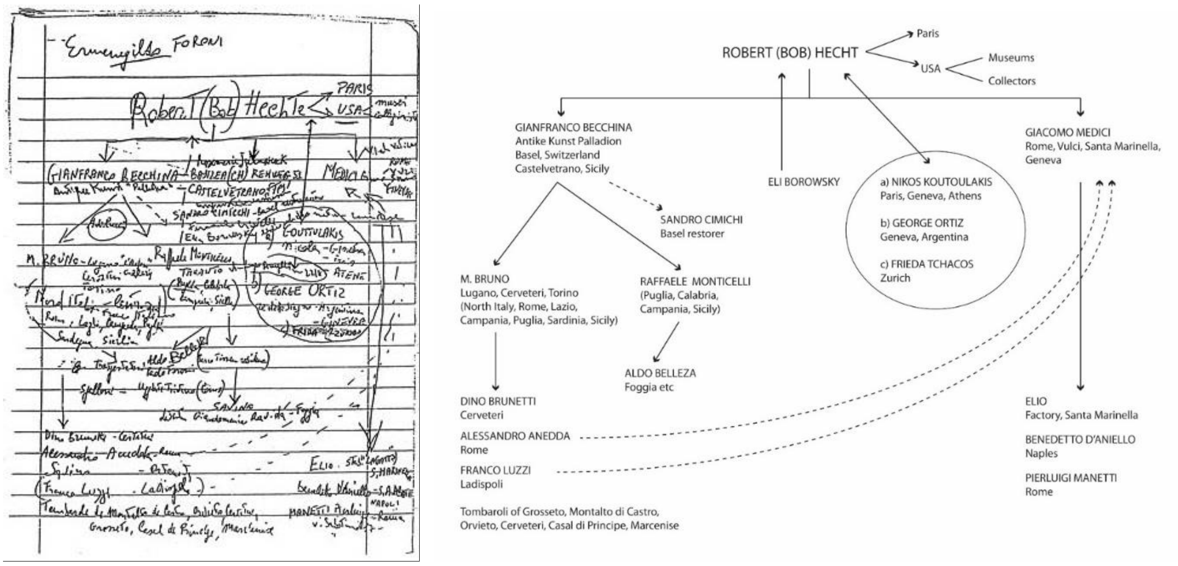

Figure 2 - Hand-drawn organizational chart showing an actual documented trafficking network, discovered by Italian Carabinieri in 1995 when raiding the premises of Daniel Zicco (Watson and Todeschini 2007), alongside a clarified version; source Neil Brodie and the Trafficking Culture website (Brodie, Organigram 2012)

stakeholders as loosely connected nodes of a network with interchangeable participants and actors (Mackenzie and Davis 2014). This is a much more complex series of relationships, but better represents the few cases of actual trafficking networks that have been uncovered through police investigations (see Figure 2).

Regardless, the hierarchical representation does provide us a useful set of terms for the roles performed at different stages of the market: **extractors**, **early-stage intermediaries**, **late-stage intermediaries**, and **buyers** (see Figure 3). The motivations and actions of these roles are primarily driven by opportunity, guided by the traditional market motivators of supply and demand (Polk 2014). In his analysis of the trade as an international criminal network, Campbell argues that these positions develop organically due to the specialist knowledge necessary to profit from the trade (Campbell 2013). Role-defining skills include the ability to locate sites, extracting and transporting material, the capability of smuggling objects across national borders, laundering and obfuscating the provenance of material, and an academic understanding of both art and history (along with associated contacts) in order to locate buyers. It is within these final two specialization areas that one most often finds individuals involved with the supposedly legal and open trade in antiquities overlapping with the illicit art markets.

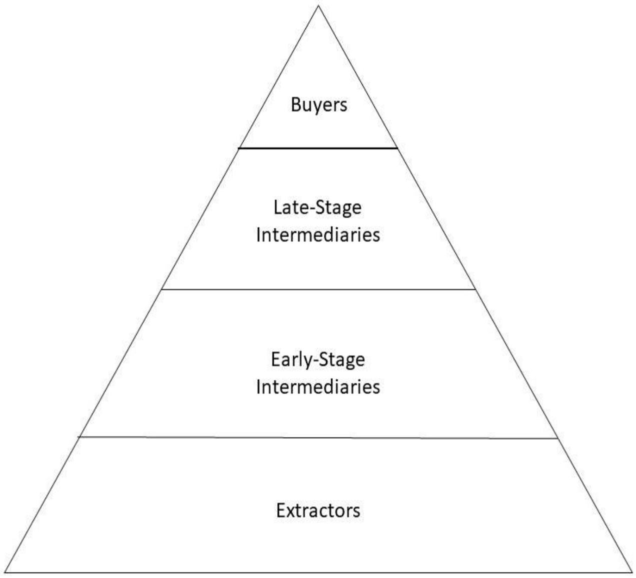

Figure 3 - Notional pyramid hierarchy of a trafficking network, source Author

2.2.1 Extractors

At bottom end of the market are the individuals acquiring the cultural material. Several terms have been used to describe this group: looter, grave-robber, and subsistence digger are some of the most often written about, with varying associated connotations (Kaiser 1991). Broadly, the objects they provide to the market were acquired in a way which has been defined as illegal. The specifics of this illegality can be broadly defined, but in general includes the breach of national laws regarding the excavation and export of archaeological material, art works and objects stolen from museums and private collections, and material objects and statuary illegally removed from public locations and cultural institutions (such as churches, temples, and other sites containing decorative elements) (Massy 2008). In their analysis of the trade, Brodie and Tijhuis agree that it is this stage which finds the most academic investigation, with both archaeological and anthropological case studies written about motivations and methods (Brodie, Uncovering the Antiquities Market 2012) (Tijhuis 2009). Returning to Campbell's analysis, he further notes that it is this end of the market that will be most varied in terms of structure and skills, based on local in the source country conditions (such as geographical, governmental, and cultural factors). A *nighthawk* performing illegal metal-detecting in rural England employs different methods from a *tombaroli* searching for and breaking into ancient tombs in Italy, both of whom have very different motivations when compared to Afghani farmers living in a combat zone or

Cambodian villagers pressganged by local military authorities. Many justifications exist for their actions at this stage: some see if as a necessary means for survival, others in less dire circumstances view it as a leisure activity or a traditional practice, and some perceive it as resistance to foreign archaeologists just as guilty of theft save with a veneer of legitimacy due to academic background (Kersel 2007) (Kaiser 1991).

2.2.2 Early-Stage Intermediaries

This is the first point of go-between from a source country to an international market, encompassing operators who specialize in moving the material beyond the local communities and provide the lowest-tier of funds for organizing the trade. In regards to the global supply chain, there is little academic documentation regarding how cultural objects actually move from the ground to the international market. Mackenzie and Davis specifically lament that nothing "approximates the sort of global trafficking research that is emerging in anthropology, history or international relations, or in criminology for other international trafficking problems, and so we are left with something of a black hole in our understanding of illicit antiquities trafficking networks" (Mackenzie and Davis 2014). Notably, the secretive nature of this stage also greatly enables the introduction of fakes and forgeries into the trade pipeline. While material cultural heritage is a finite resource, the clandestine nature of the market creates an opportunity for those with the necessary skills and connections to instead manufacture high-demand items for the market. The level of technical capability and craftsmanship exhibited by these producers is impressive, such that even expert dealers and scientists will have difficulty identifying what is a genuine artefact and what is a fake (Massy 2008). In this manner false and unproven provenance works against even other later stage actors in the market, turning antiquities into a developed product in the same way that narcotics are cultivated and illegal arms manufactured (Campbell 2013).

2.2.3 Late-Stage Intermediaries

Late-Stage Intermediaries represent the point at which laundered items transition to the legitimate trade of destination countries. Roles here are most often associated with the art market, including brokers, valuers, dealers, and auction houses (Massy 2008). Many times single individuals can occupy multiple role categories, for example providing valuation services in addition to direct sales. Some of these roles necessarily operate in secret due to the nature of market confidentiality: Massy notes that brokers in particular are often used as fronts for collectors who do not want their interests in a particular item publicized (and thus reducing the potential for a price increase tailored to their personal interest). Valuers are also central roles which are yet unregulated and generally unobtrusive, consisting of academics and art historians, dealers, and artists (or their heirs) who can claim authority in the appraisal of market goods; the rigor by which they evaluate prices for the market depends as much on the buyer as on the goods themselves. Finally, there are the dealers and auction houses which perform the actual sales. National legislation determines the particulars of how transparent or well-documented these groups must be in their sales role, and most market centers are located in highly developed economies (New York, London, Paris) as well at loosely regulated transit ports (Switzerland, Hong Kong) (Polk 2014). Even this stage can have multiple additional layers, as allied auction houses trading material can be used to create a false genealogy for the material, or to allow for manipulation of market demand. As with the early-stage intermediaries, little academic published work exists to quantify and document this side of the market. Instead, most of the work has been through media investigations and traditional journalism; the most well-known examples in recent years has come from Peter Watson. His first published book focused on the auction house Sothesby's malpractice in regards to antiquities sales (Watson 1997), and is held to be directly responsible for Sotheby's announcing that it would no longer hold regular antiquities sales in London. This was followed soon after with further documentary work he performed in conjunction with the Italian Carabinieri law enforcement group, whose criminal investigations of the antiquities dealer Giacomo Medici revealed a complex (and well-documented) supply chain leading from tomb-robbers in Italy to freeports in Geneva, and ultimately into the hands of head curators at the J. Paul Getty Museum in the USA (Watson and Todeschini 2007). As with Sotheby's, the results were indictments against some of the highest levels of professionals working with cultural material, high profile resignations under a cloud of professional shame, but with many of the tangential operators still in business today. The demand still exists, and thus the suppliers continue to profit.

2.2.4 Buyer

The demand which supports all the previous stages are ultimately the buyers. As alluded to earlier, when the source is through an auction house this often includes wealthy individuals seeking prestige pieces, or institutions augmenting their collections, providing the economic root which creates the trade. As Brodie bluntly puts it: "… it is the money of the wealthy collectors which destroys culture and corrupts societies, not the activities of poor and disadvantage diggers and impoverished bureaucracies. Clearly, if this characterization is true, it will be unwelcome, and vigilance is required to ensure that the money and influence of wealthy collectors are not employed in such a sway as to divert attention away from themselves and back towards the trade or the market" (Brodie and Doole 2001). However even this has begun to change, due to the same innovations in technology which have disrupted other traditional markets. The Internet has created a new avenue for illicit antiquities to reach potential buyers, creating new markets for lower-grade objects which would not previously have been considered worth transitioning through the auction house pipeline (Brodie 2015). Unfortunately, this opens up the market to those from a much broader socio-economic background, creating even greater demand to drive the machinery of heritage theft. Just as the front-end of market can be argued to function as a broad collection of loosely connected extractors, so too has the back-end begun to splinter into a distributed network of participants spread across the globe. Such a reality has made international cooperation in targeting the trade all the more necessary.

2.3 International Efforts to Combat the Illegal Antiquities Market

2.3.1 International Police Operations

The most recent figures from the WCO Illicit Trade Reports show variation in the material that has been found to be trafficked in the past few years, with the most recent customs reporting showing an increase between 2016 and 2017, both in the numbers of seizures and the quantity of antiquities seized (World Customs Organization 2017). The report suggests that this increase was in part due to two operations conducted by law enforcement agencies in 2017: **Operation Athena** (organized by the WCO and INTERPOL) and **Operation Pandora II** (organized by the Spanish Guardia Civil and Europol). Both were conducted during the months between October and December of 2017, and involved cooperation from customs and police in over 81 countries. More than 41,000 cultural objects were seized, primarily consisting of coins, paintings/drawings, furniture, musical instruments, porcelain, archaeological and paleontological objects, books/manuscripts, and sculptures.

Previous international operations provide us with similar insights into what types of objects are being trafficked. **Operation Colosseum** was one of the first international operations documented in the WCO Trade Reports (World Customs Organization 2015, 139), and was conducted in 2013. It was coordinated between the Regional Intelligence Liaison Office for Western Europe and Italian Customs, with participation of INTERPOL, the European Anti-Fraud Office, the European Commission's Taxation and Customs Union Directorate-General, ICOM, the Italian Ministry of Cultural Heritage, and the Italian Guardia di Finanza. It resulted in 32 seizures: the most significant finds were 70 archaeological objects seized by the Greek Police in Thessalonica. They included ancient helmets, gold masks, clay statuettes, and pots originating from Ancient Greece, and were believed to be stolen from tombs in Central Macedonia (specifically from Area Pella, in Sindos). Other seizures resulting from the operation consisted primarily of coins and books, with the largest finds recovered in Switzerland, Italy, Russia, and Cyprus (Republic of Cyprus Customs and Excise Department 2019).

Operation Odysseus (World Customs Organization 2015, 139-140)was conducted in 2014, and grew out of the heightened awareness of the illicit trade resulting from Operation Colosseum. The focus of the operation was targeted at smuggling routes from conflict zones into Europe. Many of the same agencies from Colosseum participated in this operation as well. It resulted in 43 seizures, involving a total of 44,235 objects. While significant, the vast majority of this was specifically from the seizure of 42,000 coins by Bulgarian Customs. In addition to the Bulgarian coin hoard, 1, 300 coins were seized at the border of Romania, Cyprus, France, Russia, Turkey, and Uzbekistan. The second largest seizure was antiquities (403 objects reported by Turkish Customs), and the third largest was lithographs (118 reported by France). In terms of objects specifically originating from the Middle East/North Africa region, 36 archaeological items from Egypt (amphorae, pots, vases, and other ceramic objects) were seized by Spanish customs on their way to Valencia; analysis identified a combination of authentic and fake objects. Finally, Turkish Customs also seized 11 coins and two Christian manuscripts originating from Syria.

Operation Odysseus was soon followed by Operation Pandora, split into multiple phases. The first **Operation Pandora** was led by Cypriot and Spanish Police in cooperation with Europol, with support from UNESCO and INTERPOL. It took place in October and November

2016 (World Customs Organization 2016, 8), and laid the ground work for future operations. Several police officers were deployed on the spot during the primary operation week to assist national authorities with inspections and searches. INTERPOL assisted investigators in the field by cross-checking hundreds of objects against their stolen works of art database, and also provided a swift response when identifying artefacts of illicit provenance. The WCO supported the joint action by facilitating the communication, cooperation and assistance between law enforcement and concerned customs administrations. UNESCO contributed to the operation by providing training materials and offering recommendations to the participating countries. 3,561 works of art and cultural goods were seized, almost half of which were archaeological objects; 500 archaeological objects were found in Murcia, Spain, of which 19 were stolen in 2014 from the Archaeological Museum in Murcia. Over 400 coins from different periods were seized following investigations into suspicious online advertisements. In total, this operation led to the arrest of 75 individuals. In 2018 this was followed by Operation Pandora II (reportedly recovering more than 41,000 objects and leading to 51 arrests), and in 2019 Operation Pandora III (recovering 18,000 cultural objects, and leading to 59 arrests) (Anderson 2019).

2.3.2 International Documentation Standards

In addition to coordinating with international law enforcement operations, several of the organizations concerned with heritage preservation have sought to combat the illicit trade by supporting better documentation standards for archaeological material. Two of the most relevant and widely adopted standards include ObjectID (supported by ICOM and the J. Paul Getty Trust) and the UNESCO-WCO Model Export Certificate (supported, unsurprisingly, by UNESCO and the WCO).

2.3.2.1 ObjectID

ObjectID was initially released in 1999 (Thornes, et al. 2000), developed through consultation and negotiated agreement between cultural heritage stakeholders. This included museums, archaeological organizations, police and customs officials, the insurance industry, and art and antiquity professional appraisers. The primary development was done by the J. Paul Getty Trust supported by UNESCO, ICOM, the Council of Europe, and the United States Information Agency. ICOM in particular has encouraged the distribution and use of ObjectID throughout the world, in part to support efforts to combat illegal trafficking of antiquities. ICOM's support has been key to its widespread adoption, as one of the most recognized contributions by ICOM as an organization has been their leadership role in drawing attention to the ethical and legal challenges presented by the illicit antiquities trade (Nafziger 1972). These agencies now work together in encouraging the implementation of the standard, as well as developing the networks necessary to circulate the information. In doing so, they are addressing the two primary principles that had been identified as necessary to combat the illicit antiquities trade, namely:

1. a stolen object cannot be returned to its rightful owner without adequate documentation, and
2. documented information about an object should be capable of rapid, global communication amongst a number of disparate organizations.

ObjectID's development was achieved primarily through the use of international questionnaire surveys circulated amongst organizations in over eighty countries. The result of this work was an agreed upon standard of documentation intended to fulfil the essential requirement of being easy to implement and capable of being used by both specialist and non-specialists dealing with cultural antiquities. This was achieved primarily by the publication and dissemination of the document titled *Introduction to Object ID Guidelines for Making Records that Describe Art, Antiques, and Antiquities* (Thornes, et al. 2000). This guideline document is split into two sections: *Part 1: Describing Art and Antiques Using Object ID* and *Part 2: Photographing Objects for Purposes of Identification*. Part 1 describes the core of ObjectID, detailing the nine categories of information to be recorded as a minimum necessary standard of description for identifying a cultural object, and four steps necessary to fulfill the procedure:

- Nine Categories
    - Type of Object
    - Materials and Techniques
    - Measurements
    - Inscriptions & Markings
    - Distinguishing Features
    - Title

  - Subject
  - Date or Period
  - Maker
  - Short description
  - Single image
- Four Steps
  - Taking photographs of the object
  - Informing the above mentioned categories
  - Writing a short description including additional information
  - Keeping the constituted documentation in a secure place

Part 2 is a set of recommendations for photographing an object for the purposes of documentation (rather than for aesthetics or promotional usage). The main areas of advice are on choosing viewpoints for photographing the object, creating backgrounds, and setting up lighting. The main criteria for these focus areas hinges on the material and type of object, with broad and generalized suggestions. Significantly, these guidelines are written with the assumption that photographic records will be taken using traditional film, rather than as digital images. This is further reflected in the advice for lighting and background, along with the assumptions regarding imaging workflows. In terms of modern recording, there is a significant gap between current tools and the baseline assumptions around which these recording criteria were developed, unfortunately.

2.3.2.2 UNESCO-WCO Model Export Certificate

Another response seeking to address the international illicit antiquities trade was the development of a Model Export Certificate (MEC) for Cultural Objects (World Customs Organization 2005). This was a joint effort between UNESCO and the WCO, specifically intended to address the movement of cultural objects and property across national borders. The intent of the MEC is to enable the identification and traceability of a cultural object, as well as to distinguish them from ordinary objects.

- Significant sections:
  - Heading 7: Owner of Cultural Object
  - Heading 8: Photograph of Cultural Object <space for single photograph>
  - Heading 9: Dimensions and net weight of the Cultural Object

- Heading 10: Inventory number or other identification
- Heading 11: Description of Cultural Object
- Heading 14: Materials and Techniques

The MEC shares many similarities with ObjectID in the textual descriptions. However, it is a physical form and template, which gives it some limitations. Significantly, for the photo section there is only room for a single documentation photo; thus the customs official is given only a single perspective view of the object (and is susceptible to traffickers using perspective and lighting to minimize identifying features). In many ways both standards are rooted in assumptions about the difficulty in capturing a visual record of an object, and have not yet been updated to better reflect modern digital capture and record keeping.

2.3.3 ICOM Red Lists

In addition to the above two standards, one of the primary documentation resources regarding trafficked object types are the publications of Emergency Red Lists by ICOM (International Council of Museums 2019). ICOM has been creating and distributing these documents since 2000; their purpose is to aid customs officials in the identification of illicit antiquities and cultural material. They are not lists of actual stolen objects, but are instead categories or types of objects most at-risk from specific regions, with visual examples. The lists are distributed as printed booklets, but are also available as digital downloads from ICOM's website. ICOM has also added an ability to on their website to search for types of objects delimited by *Material*, *Type of Object*, and *Country*[4]. Currently, booklets have been created for regions in Asia (specifically Afghanistan, Cambodia, and China), Latin America and the Caribbean (Central America & Mexico, Colombia, Dominican Republic, Haiti, Peru, and general Latin American antiquities), the Middle East and North Africa (Iraq, Libya, Syria, Yemen), and West Africa (Egypt and general West African Antiquities). These lists compose the most comprehensive catalogue of the types of artefacts supplied through the illicit antiquities market, and the object categories which customs officials and heritage collection organizations much be most vigilant for as potentially looted material.

[4] https://icom.museum/en/resources/red-lists/

2.4 Conclusion

"Ultimately, the looting of cultural material will only stop when collectors, museums and dealers refuse to buy unprovenanced objects. No matter what protective measures are put in place, whether draconian or liberal, they will be circumvented if a demand is created by a purchaser with few scruples or principles." (Brodie, Doole and Watson, Stealing History: The Illicit Trade in Cultural Material 2000)

Decades of investigation by heritage professionals and international bodies seeking to protect cultural patrimony has, unfortunately, primarily shown that the illegal antiquities market is a persistent and ongoing threat with no obvious set of perfect solutions. We know the illicit antiquities trade is poorly documented when compared to other types of illicit markets. International action tends to result during periods in which well-publicized threats to cultural heritage are in the public conscious, though the long term results of those actions is uncertain due to lack of data. The successes by law enforcement and customs officials mostly hints at the extent of the problem, showing that the market is both widely distributed and complex, often involving multiple individuals and groups working both in collaboration and in competition with one another. Further, the trade is not strictly hierarchical, though there are patterns of common roles in the way it functions. Significantly, many of the primary nodes on the network blend together both legal and illegal trading, linking the legitimate and black art and antiquities markets together. Often this is done in later stages of the trafficking chain, with specialists who falsify provenance and trading records to create a false history. Heritage organizations and international law enforcement groups have supported the wide-spread agreement and adoption of documentation standards necessary to accurately categorize and identify objects moving across international borders and into museum collections, through both visual and textual labelling. Unfortunately, antiquities change hands multiple times along the artefact trafficking chain, and the use of forged and altered photo documentation is a common tactic for bypassing customs and disguising the nature of a looted object. Moreover, the traditional source of demand which fuels the trade (auction houses and private collectors) has itself begun to change, thanks to changes in technology opening up trafficking networks to new potential buyers. Despite international operations acting directly against the trade itself, it is becoming increasingly evident that heritage protection agencies must also focus on improving and enhancing the tools and information sharing networks used to combat trafficking. Consistently the most necessary requirements identified are 1) an appropriate level of data recording for an object to create a provenance trail, and 2) a method by which a non-expert can rapidly identify an object which may have been trafficked (specifically against attempts to disguise its identifying features and characteristics). Thus any potential solutions we wish to investigate in our research should use these as the guiding principles by which our software's effectiveness should be measured and judged. Having identified the nature of the problem, it is now possible to consider potential solutions based on the data at hand.

3 DATA UNDERSTANDING AND PREPARATION

3.1 THE ORIENTAL MUSEUM'S EGYPTOLOGY COLLECTION

The Durham University Oriental Museum is an academic institution originally created to house the University's growing collection of artefacts for the School of Oriental Studies in the 1950s. Its purpose was to support research and teaching at Durham University, and thanks to several important acquisitions early in its history it is now in possession of over 30,000 artefacts from several near and far-east cultures. The Egyptian and Chinese collection has been granted ''Designated Outstanding Collection'' status by the Arts Council England, and is considered one of the best Egyptology resources in Britain (Arts Council England 2014). According to their collections database most of this material comes from two sources: the Northumberland Collection (roughly 2000 items) and the Sir Henry Wellcome Collection (over 4000 items).

3.1.1 The Northumberland Collection

The Northumberland Collection is the Oriental Museum's foundational collection, its acquisition laying the foundation from which the rest of the museum was developed (Barclay and Barclay 2018). This work was overseen by Professor Thomas W. Thacker, a staff member and eventual Director of Durham University's School of Oriental Studies. He held the opinion that in order to properly understand and research ancient cultures and languages it was vital to have access to their material culture. Given this view, when the 10th Duke of Northumberland sought to relinquish his family's collection of Egyptian and Near Eastern antiquities Professor Thacker began organizing a campaign to bring them to Durham. In this effort he was competing against the British Museum (where the collection had previously been stored on-loan) and the Brooklyn Museum in the USA. In making his case, Thacker leaned heavily on arguments of preservation and accessibility. In his view the collection would be diminished within the British Museum's archives, blending in with the other Egyptian material there. Further, the experiences of targeted bombardment from the Second World War had emphasized the danger of placing too many irreplaceable antiquities in any one location. In his words: "In these days of aerial warfare egyptological collections are not immune from destruction, and what is destroyed can never be replaced. Because of the menace from the air it is unwise to allow our treasures to become too centralized or always deposit them in towns which could become military targets." (Thacker 1949) Considerations of preservation and risk from conflict were at the root of the Oriental Museum's collections from the start, though in a different context (global vs regional conflict) and set of ethical considerations (heritage preservation for British archaeological institutions and practitioners rather than the Egyptian inheritors). Ultimately Professor Thacker was successful in convincing the Duke of Northumberland that the collection should remain in the North East of England, and after a payment of £12,000 it was delivered to Durham University in the summer of 1950. Initially referred to as the Alnwick Collection, it eventually was given its current records label of the Northumberland Collection. This was in part due to the British Museum period of ownership, with accession numbers included an 'N' or 'North' prefix that is still found in the management system today (as well as inscribed on some of the objects). It also derives from the correspondence of the collection's initial cataloguer, Samuel Birch. It is from Birch that much of the details about the early development of the collection derive, specifically the parts of the collection assembled by Algernon Percy, the 4th Duke of Northumberland (Birch 1880).

3.1.1.1 Algernon Percy (Lord Prudhoe)

Algernon Percy was not originally intended to succeed his father as Duke of Northumberland, born as the second son to the second Duke in 1792. Instead he came to his position in 1847 after the unexpected death of his elder brother; by that time he had already established an interest in Egypt, having taken part in several journeys through the region between 1826 and 1829 (Ruffle 1998). His journals document travels with Major Orlando Felix and highlight Lord Prudhoe's interest in Egyptian history, as well as his academic familiarity with hieroglyphic scripts. Though they record his experiences with well-known sites of the time, there is little to indicate that the Duke acquired much Egyptian material during these specific travels. He does briefly mention a pair of lions at Jabal Barkal now known as the Prudhoe Lions, which are currently on display in the British Museum (see Figure 4). However, his journals do not describe the circumstances of their acquisition and removal. They do, however, contain Lord Prudhoe's awareness of known monument deterioration in the country, as well as an acknowledgement of damage to the country's heritage through the looting of antiquities. Nevertheless, it was precisely the antiquities trade which became the primary source of Lord Prudhoe's collection (Barclay and Barclay

2018). Throughout the 1830's he acquired the bulk of his material through London auction houses (including a large lot from Henry Salt in 1835), apparently much influenced by the advice of the Egyptologist Sir John Gardner Wilkinson (who dedicated his seminal work *Manners and Customs of the Ancient Egyptians* to the Duke's elder brother, Hugh) (Wilkinson 1837). Over this period and up until the Duke's death in 1865, the collection grew from around 160 Egyptian items to an assemblage containing over 2000 objects (Armstrong 2018).

Figure 4 - The Prudhoe Lions, source British Museum (Trustees of the British Museum 1835)

3.1.1.2 Henry Algernon George Percy, Earl Percy (Lord Warkworth)

The collection would further expand nearly half-a-century later, thanks to the efforts of Lord Prudhoe's inheritor, Henry Algernon George Percy (Earl Percy, also sometimes referred to as Lord Warkworth). It was his father (the 7th Duke of Northumberland) who employed Samuel Birch to catalogue the collection for publication. Lord Warkworth himself produced two travelogues detailing his exploration of Ottoman Turkey: Notes of a Diary in Asiatic Turkey (1898) and Highlands of Asiatic Turkey (1901) (Warkworth 1898) (Percy 1901). From these volumes we know he had a desire for travel in search of ancient Near East civilizations and their monuments; his primary concern was with Assyrian and Hittite cultures, though he used the two terms interchangeably (as well as referring to most any ancient writing as "hieroglyphics"). He documented several instances wherein he acquired archaeological objects while travelling in this region, most often purchased from local peasants or specialist dealers (Warkworth 1898, 201-207, 215, 231, 261). Additionally, his engagement with academic and exploratory groups and his eventual role as Under Secretary of India (1902-1903) and then Foreign Affairs (1903 – 1905) would have given him access to antiquity markets (both in London as well as abroad). The full extent of his contribution to the collection has been unclear, in part due to the collections transition through the hands of several owners and institutions, with uncertainty introduced through a succession of accession registers, card catalogues and various computer databases. The Oriental Museum has in recent years sought to compare the contents of the collection to various historical records; from their work, they believe that Lord Warkworth's contribution likely amounted to roughly 178 Egyptian antiquities (primarily Scarabs, beads, and seals) (Barclay and Barclay 2018).

Though the Northumberland Collection does display some of the characteristics of antiquities acquired through the illicit antiquities market (obscure provenance records as well as direct purchase from Extractors), by and large the material is reasonably well documented in published catalogues written by Victorian contemporaries of the Percy's, as well as from time spent in the archives of the British Museum. Instead, it is the other branch of the Egyptology Collection which provides a material background more closely matched to our intended area of research into artefacts of dubious provenance: the Wellcome Collection.

3.1.2 The Wellcome Collection

While the historical details of the Northumberland Collection has been researched to an extent, the background and acquisitions of Sir Henry Wellcome's Collection has been subject to much more extensive documentation and study (Russell 1987) (Tansey 2002) (Engel 2017). Specifically, the biography *An Infinity of Things* by Frances Larson serves as perhaps the most in-depth and revealing insight into both Wellcome as a collector and the nature of the collection he built (Larson 2009). It is from this work that we have built the greatest understanding of just how Wellcome created his own supply network of antiquities.

Sir Henry Wellcome was a businessman originally from Almond, Wisconsin. His early career focused on the area of medicine and pharmaceuticals, and in 1880 he established a partnership with a fellow student from the Philadelphia College of Pharmacy, Silas Manville Burroughs. Together they began a drug and medical production company in London, and over the next decade the Burroughs, Wellcome and Company managed to become one of the most successful pharmaceutical companies of the late nineteenth century (Tansey 2002).

According to autobiographical accounts, Henry Wellcome had first become interested in collecting as a child, inspired by native arrowheads found in his home country. Wellcome's collecting began in earnest after the death of his business partner, when Wellcome became the sole proprietor of the pharmaceutical company. Soon afterwards he placed greater emphasis and resources towards privately funded research, opening several laboratories scattered across the world (including a lab in the Sudan capital of Khartoum in 1902 and a floating laboratory on a boat located upon the Nile in 1905). This would lead to pioneering research in the medical industry, but also allowed Wellcome to begin pursuing his own personal research interests into the history of medicine (Larson 2009, 25-27). It was while Wellcome was surveying the needs for his Khartoum research laboratory that his fascination in the local region's material culture and history was established. Acquisition of artefacts and material from regions in the Middle East and Africa was conducted with the same organizational focus as his other business practices, with the company's departments tasked with classifying, recording, and storing material Wellcome acquired on his travels (Larson 2009, 32-49). By 1901 many items from of his collection had been put on display in England, including a wide variety of modern Egyptian tents and couches, as well as weapons and objects from antiquity used as decorative displays in the companies communal areas. This newfound interest and set of acquisitions led him to consider a new area of scholarship. He began introducing himself to scholars and experts of the region by gifting some of the material he had collected. A decade later this interest would lead Wellcome to return to the Anglo-Egypt Sudan region, this time focusing on an excavation in the hills of Jebel Moya (located in the middle of the Nile basin). Between 1910 and 1914 he oversaw numerous excavations here, and was only given pause by the outbreak of World War 1. Though he boasted that this location would lead to the original birthplace of human civilization, it would turn out that the site mostly consisted of a late Neolithic settlement. Nevertheless, he excavated a massive amount of material, shipping back ancient stones, bones, and pottery fragments, insisting that every discovered scrap must be retained, classified, and labelled (Larson 2009, 58-59). Still, his acquisition of antiquities through site excavations would soon pale in comparison to his primary method of collecting. As time went on, the vast majority of his acquisitions came via the auction markets of early 20th century London.

3.1.2.1 The Wellcome Collection and the Antiquities Trade

"Although the number of public museums grew enormously during the nineteenth century, the auction houses had long provided a forum for connoisseurs to study the latest offerings on the market, and of spectators to marvel at rare artefacts. As the century progressed, the salesrooms filtered a profusion of exotic specimens brought home by surveyors, builders, government officials, military men, missionaries, and medics stationed overseas." (Larson 2009, 78-79)

Wellcome's first major purchase of books at auction occurred at Sotheby's, in December 1898 (Larson 2009, 19), the beginning of what would be a long relationship with the antiquities trade. It was through auctions from throughout England that the Wellcome collecting truly took shape. While Wellcome would involve himself personally while in London, much of this was also done through a team of agents hired employed solely for the purpose of acquiring archaeological material. The market was often controlled by groups of dealers who would collaborate (or conspire) with other similar specialists in order to maintain a measure of control over their specialization within the trade. Most private collectors chose to collaborate with these sorts of dealers rather than attempt to compete against them. Wellcome, however, chose not to commission dealers in auction rooms: "Dealers regularly sent Wellcome material privately, on approval, but he preferred to pit his tactical skills against them at sales." (Larson 2009, 81) Moreover, he insisted on secrecy on the part of his agents, such that dealers and other collectors should not be aware of what he was after. However, the sheer amount of material he was acquiring meant that his agents became known to the dealers as having access to a very wealthy provider. His personal fortune created an entire branch of a Late-Stage Intermediary network within the antiquities trade of the time. Unlike other contemporary collectors, though, Wellcome was not seeking to create a collection of expensive prestige pieces to display his wealth. He sought instead to acquire items at low cost, desiring a collection of practical and encyclopaedic display of material culture. He desired commonplace items, and duplicates of those items whenever available; his vast sums of money translated into vast numbers of objects. This was not an undirected hoard of dealer's junk, but a meticulous effort in his sense of academic pursuit. He considered collecting as his contribution to the academic tradition he wished to be considered a part of, and his collection was intended to be practical (contributing to corporate designs and medical research). Further, he took inspiration from the educational museums of the late nineteenth century, specifically calling out the Pitt Rivers Museum at Oxford and the Horniman Museum at Forest Hill as the nearest approximation of his own aspirational museum. These contrasted with the Cabinets of Curiosities of his time, whose intended purpose was wonder and oddity. Instead Wellcome's intent was to form a scientific, anthropological collection of common and everyday objects. His records soon began to map relationships between how objects and their makers were to existing counterparts, becoming much more of a classification exercise. This required large numbers of objects of similar type, with which structured comparisons and classification could be made. (Larson 2009, 85-89) "It became important to gather together as many varied objects of the same general type as possible, to ensure that the resulting picture of the world was through. If everyday objects were a kind of historical data, then collections were giant data sets, and missing objects meant missing data, which could very well lead to faulty conclusions regarding the human past." (Larson 2009, 89)

Despite the academic aspirations, artefact acquisition was also a source of personal validation and excitement for Wellcome. One of his earliest agent's was Charles John Samuel Thompson, employed in a position that was officially designated as "Librarian." Thompson's duties in this regard were varied, travelling across the UK at Wellcome's request to seek out all manner of books and manuscripts. Moreover, he acted as a historical consultant and archaeological researcher for Wellcome, often providing the background information used by the company's advertising division. It was his work in the trade auctions, however, that most held Wellcome's attention: his weekly reports had been a great source of entertainment to the pharmaceutical magnate. Thompson had enjoyed his escapades in the salesrooms and his dealings with fellow collectors, and he described his collecting adventures with assurance and style. He knew that Wellcome loved to read about his work as though it were an adventure, in which scheming rivals were duped and Wellcome's staff returned home heroically, bargain prize in hand (Larson 2009, 203).

Although Wellcome was not as consumed with unique prestige pieces as were his wealthy contemporaries, this was still very much about trophy hunting for him. For Wellcome, the overriding goal was acquisition, an attempt to physically house and display material culture in order to finally understand how it all interrelated. To this effort he created an organization dedicated to succeeding in the marketplace of the antiquities trade, in order to secure his place in the adjacent academic community. Wellcome had created a collection business inspired along similar lines as his business operations, but instead of profit it was given to the purpose of collecting. Overseen and directed by Wellcome as he would any business venture, he was directly involved in his agent's actions, insisting on monthly updates and material results. "[his agent's] grievances regarding report writing and payment schedules in the late 1890s can be seen as a clash of intellectual cultures, between the scholar who believed that historical research should be truly collaborative and qualitatively judged, and the businessman who required accountability and quantitative results." (Larson 2009, 115) Moreover, his direction to his agents once more displayed his collecting worldview, prioritizing breadth and number above all else: "Thompson was constantly relaying Wellcome's insistence that 'nothing should be overlooked, and no part of the country skipped' in the hunt for antiquities. … He must 'leave no stone unturned' and should not come home until 'India is completely ransacked as far as we possibly can for literature and other objects of interest connected with ancient medicine, and all the great centres of learning, visited and ransacked'." (Larson 2009, 119) He was also insistent that all work they did was ultimately in service to his own vision, relying heavily on non-disclosure agreements and contracts stating that all work done by his agents be attributable to himself. Moreover, they were disallowed from pursuing their own interests in areas which Wellcome believed would put them in competition of his own academic aspirations. Personal collections were disallowed, and personal research discouraged. Ultimately, his treatment of his agents often resulted in soured relationships and bitterness. (Larson 2009, 113-126)

3.1.2.2 The Undocumented Wellcome Collection

Though Sir Henry said his ultimate aim was to create a museum 'as an institution of post-graduate study', he never opened his collection to a wider set of researchers. His staff were discouraged from seeking publications or presenting their work at conferences, and discussion of the collection with anyone outside the institution was forbidden. This contrasted sharply with the company's Scientific Research Bureau, which published a total of 430 academic papers between 1897 and 1921, as well as being encouraged to participate in academic conferences and engage with fellow scientists. (Larson 2009, 180-181) World War 1 had a significant effect as well, both reducing the number of skilled staff from the museum (sent out to support the war effort) as well as creating a new output of antiquities coming to the auction markets at very attractive prices. "As Wellcome systematically amassed rare historical documents and relics from all over the world, the wider scientific community began to take note, but their requests for information were answered in the vaguest terms and questions from visitors were evaded." Further, "Wellcome was simply unable to share his collection with anyone else until he deemed it presentable, but there was no hope of this while his acquisitive urges continued to run unchecked." (Larson 2009, 182)

In essence, Wellcome had allowed his need for completeness and secrecy to overrule his stated goals of scientific research. This was exacerbated by the antiquities market he had sought to master: "His main priority was still acquiring artefacts and books, and, while this was the case, he feared high-profile research work would compromise his tactics as a buyer by prejudicing dealers." (Larson 2009, 192) He stated as much in writing in 1928: "In principle our policy might be expressed in the words 'We will say what we are going to do after we have done it'. I see no good reason for informing our would-be rivals or anyone else in advance or at any time, when, where, or how we obtain our materials." (Larson 2009, 192) Accordingly, he kept photo documentation limited as well: "He was reluctant to allow any photographs of the Museum to appear in the press, and insisted that all images had to have the appropriate copyright inscription 'being prominently imprinted on the negatives so that the words cannot be eliminated when the photo is processed for illustrations for our own and outside publications'. His strict rules often meant that national publications, particularly newspapers, refused to publish illustrations of the Museum at all. Even if the London editors had agreed to print them, images would still be cropped or altered by syndicated papers." (Larson 2009, 193) While Larson argues that Wellcome was uninterested in publicity, it also contains a quote in which Wellcome indicates an opinion that photographs diminish his collection: "… pictures are much more beneficial to journals because of their attractiveness to their readers while to us the more we publish pictures of objects in the Museum we make them commonplace and many people are satisfied to look at the picture and save themselves the trouble of going to the Museum. I hold very strong and definite views on this point." (Larson 2009, 193) At one point a newly hired head archivist (Louis Malcolm) sought to change this stance, but to no avail: "Malcolm tried to persuade Wellcome to loosen his copyright restrictions so that more photos of the Museum could be published, but Wellcome was unmoved, claiming that photographs made the Museum seem 'common-place' and actually dissuaded people from visiting." (Larson 2009, 202) The collection at that point had become sprawling, with artefacts kept in eight different storage locations spread across London. Packing crates were stacked haphazardly: sometimes upside-down, sometimes with their labels hidden. Narrow corridors and gangways were left between the various packing crates, and even those navigation alleys came to be filled, eventually.

The result was a collection in disarray, where no one knew precisely what was in storage. "Cases remained unopened for decades. Many had not been examined since they were first delivered from the salesrooms. In 1927, as the staff began reorganizing the stores, they found themselves opening cases untouched since 1905, when the Museum had been in its infancy and Thompson had led a small team of local buyers." (Larson 2009, 247) Worse, the facilities themselves were not protected from conservation issues such as moisture and dampness: textiles rotted, photo negatives stuck together, pictures molded over, and small arms rusted. This problem was fully understood by Malcolm, but could not be practically solved due to the lack of space. Moreover, his suggestion that large collections not be purchased while they searched for adequate storage facilities for the current collection was brushed aside. Scientific staff were employed simply to document long forgotten antiquities, though in many cases their work was simply to unpack objects, catalogue them, the repackage them again. "Theodore Gaster recalled, 'the constant thrill of recovering some of the most precious antiquities from the Ancient Near East which Wellcome had purchased years ago and which no one had heard of since'." (Larson 2009, 203)

It was only in the final years of Sir Wellcome's life that his collection began to be documented, when in 1935 Louis Malcolm had left and a new head archivist (Peter

Johnston-Saint) took over. He implemented new systems for registering and re-registering the vast storehouses of objects, developed based on consultation with colleagues and collectors in other museums. Moreover, this push included a renewed focus on conservation of material, not only documenting the collection but also focusing on effecting repairs to damage incurred during their period of storage. Though he sought to communicate these plans to Sir Wellcome in an effort to help his employer achieve his goal of creating a museum displaying the decades of industrial acquisition, in the end it was too late. Sir Henry Wellcome died on the 25th of July, 1936, leaving behind vast storehouses of undocumented and hidden artefacts taken from around the world as his legacy.

3.1.2.3 Oriental Museum Acquisition of Wellcome Trust

After his death, control of his assets and business (grouped together into the Wellcome Foundation Ltd private company) fell to a group of Trustees; this included handling all of the archaeological material Wellcome had collected over his lifetime's pursuit. In practice, this led to the fragmentation of Wellcome's collection. Without Wellcome's personal enthusiasm nor a functioning museum in place to curate his legacy, those overseeing his estate were unable to justify maintaining the chaotic storage. "Wellcome's trustees were faced with thousands of unopened packing cases, many filled with artefacts of unknown provenance and quality. Now that Wellcome's vision could no longer hold this largely impenetrable collection together, practical and financial considerations took precedence for the first time and dictated that the collection must be rationalized." (Larson 2009, 271)

On 16 April, 1949 an agreement was reached with the British Museum to receive thirteen hundred cases of ethnographic material for storage, with first choice regarding the dispersal of their contents (Russell 1987). Over the course of the next five this vast store was broken up and distributed to museum's both within and outside the UK. However, the disposal of the Egyptology collection was actually organized by the Petrie Museum, first in 1964 (primarily focusing on material excavated by Flinders Petrie), and then in 1971; it was this 1971 distribution that resulted in the Wellcome Material finding its way into the Oriental Museum's archives (Russell 1987). Representatives from museums with large Egyptology collections (such as Birmingham, Liverpool, and Swansea in addition to the Oriental Museum's predecessor museum, the Gulbenkian Museum of Oriental Art and Archaeology) were invited to acquire cases of material to be loaned to their museums. In total there were roughly three hundred crates of material, and as such a great majority of the antiquities acquired in this way were not chosen individually, but simply as a bulk package of goods. In terms of documentation, as best can be determined Birmingham received the majority of accession cards associated with the Egyptian material, with some few retained by the Petrie Museum.

For the Oriental Museum, it has been estimated that number of Egyptian objects which entered into their collection from this Wellcome distribution was around 4000 individual items, though even today they are still working to match historical accession register information with existing archival and database sources (Engel 2017). We thus have a collection of material originally acquired prior to the UNESCO 1970 convention, with minimal provenance information, and which was likely acquired through purchase from the antiquities trade (and thus fitting the intended focus of our research).

3.2 The Oriental Museum's Digital Records

For our own research, we sought to experiment with object classification by using a modern CNN architecture as applied to the digital records of a heritage institution. To differentiate our image set from previous efforts, we focused our attention towards digital imagery containing multiple perspectives of structurally similar three-dimensional artefacts. Our intended goal was to test its utility in classifying broad cultural strands and typologies using open-source deep learning software packages and models. Further, we sought to investigate the resulting high-dimensional data using popular data visualization packages. In order to achieve these aims, we initiated our work utilising building off a pre-established dataset created previously from the image records of the Oriental Museum.

Since 2001 the museum has been developing their digital records of the collection using a workflow that emphasizes consistency in taking perspective shots for each artefact, capturing front, back, left, right, top and bottom views (as well as detail and quarter shots when deemed appropriate). Photos are systematically ordered and labelled according to View Codes and Accession Number (a unique identifier based on when and how the item was acquired). Internally the museum maintains control of its collection using the AdLib Collection Management System (CMS), a Custom-off-the-Shelf (COTS) product from the Axiell Group built on top of a Microsoft SQL database architecture. Each object has an entry in this system which includes object categories of information; these categories were chosen to comply with the Spectrum 5.0 UK collection standard (Collections Trust 2017). The Spectrum 5.0 standard has information requirements specifically focused on metadata associated with the long-term maintenance of a museum collection, and is one of the standards to which ObjectID and CIDOC were developed to conform to (ICOM 2012). This system is located on an internal network, and restricted to museum staff (with levels of permission and user groups controlled by the software). Object images are limited to thumbnails of size 300x300 pixels, as the system does not handle full-scale images well; the collections staff further suggest that 20 to 30 thumbnail images are the upper limit for an individual record. The system has a linked thesaurus with control terms, and associates related metadata under tabbed subject groups, which can be exported to a spreadsheet data document. The museum makes provides open accessibility to the collection through a front-end webportal called Discover Durham Collections[5], which unifies many of the electronic resources for Durham University (including e-journals, databases, University theses, image resources and research datasets). The information provided through the Discover website is an abridged subset of what is contained within the AdLib system, primarily consisting of Title, Creation Date, Description, Subject, Dimensions, Physical Description, Material, Object [Accession] number, Production place, Production period, and the associated thumbnail images.

3.2.1 Initial Research – Machine Learning vs Traditional Feature Extraction

When an initial version of the dataset was created in Summer 2017, the focus was on two broad cultural strands of object: Egyptian and Sinosphere (mostly objects from China, but also including collections from Japan and Korea). The work was in part inspired by tours of the museum in which students noted interesting similarities in the objects contained within the China and Egypt galleries. The primary goal for this research was to develop a large enough collection of associated images to compare the performance of a neural network machine learning system to a traditional feature extraction recognition algorithm. We believed that the advantages of utilizing this image data with a CNN is that it avoids many of the issues encountered when trying to generalize learning towards a more “natural” image collection (Rajalingham, et al. 2018); the uniformity of data especially lends itself well to the

[5] https://discover.durham.ac.uk/

categorization process. The initial work was in gaining familiarity with the data, determining what would and would not work. Full scale images were provided by the Oriental Museum, which needed to be scaled down to the CNN input size (224 x 224 pixels). One initial problem was the presence of images with rulers and color cards; it was felt that using this data would cause the classifier to train on the rulers, rather than the objects themselves. A secondary issue was mixed data; the delivered image files included all digital records used by the museum, including promotional material and files named for a specific exhibition or special project. Duplicate data resulted from this, including identical images with different file names. It also included objects not actually in the museum's collection, but simply on loan for a short period. Nevertheless, a dataset was developed based primarily on figurine and vessel images, which was run on a CNN previously developed by the Durham University Innovative Computer Group.

While the results were mixed (only achieving at most a 94% validation accuracy with non-convergent loss and optimization), we did find that the data itself showed good clustering and type correspondences (for example, clustering a Heart Scarab amulet with other scarab amulets and impression seals, despite having been mislabeled as Anthropomorphic due to the presence of a human face). Also, the comparison testing showed better results from the CNN when compared to the traditional feature classifier. Finally, this effort served as the first iteration of data familiarization with the Oriental Museum image files, proving useful for further future revisions.

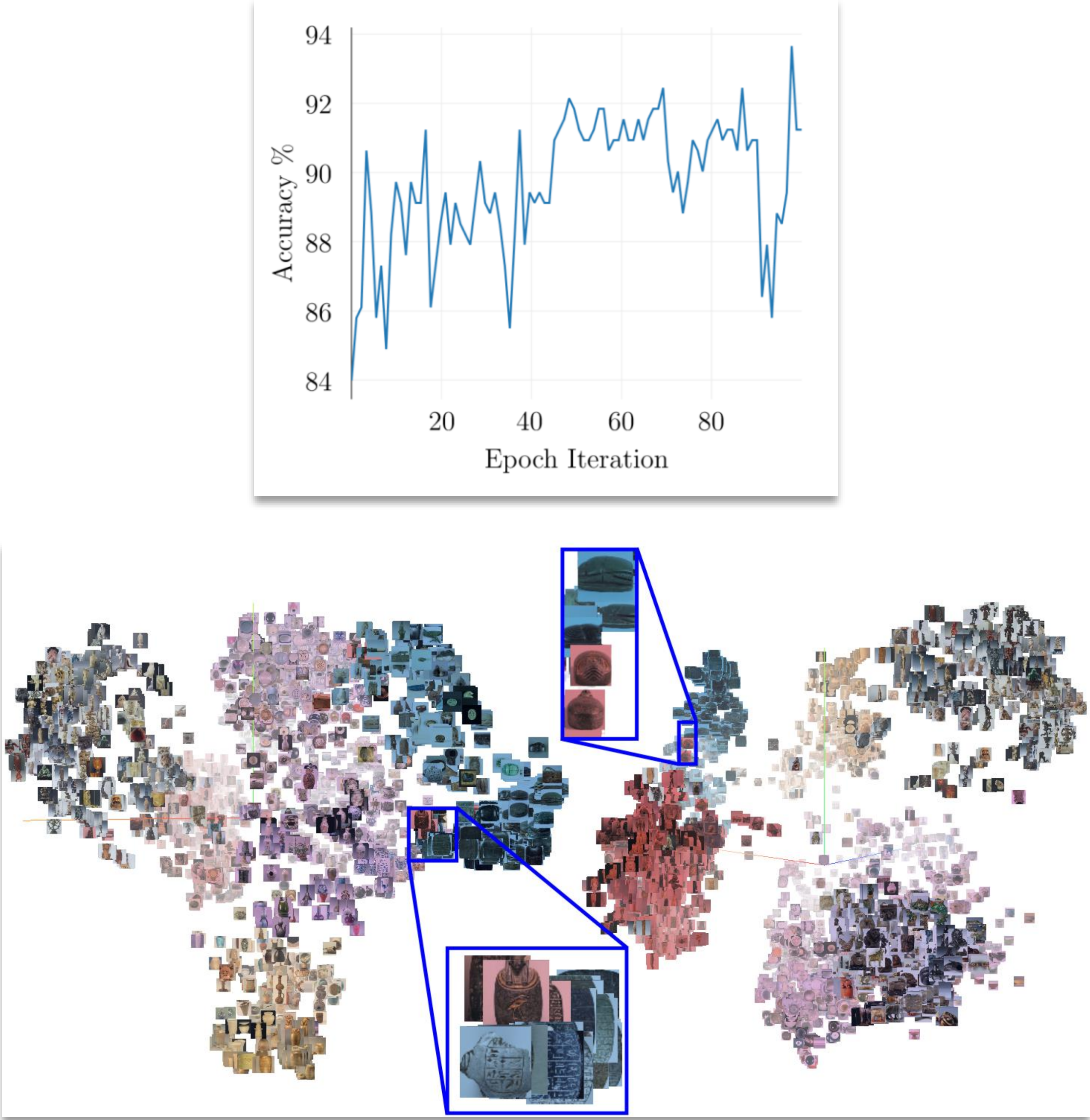

Figure 5 - Non-Convergent Validation Accuracy and photo clustering using t-SNE; source Author

3.2.2 Egyptian Data Set Development

To build on the previous research, our initial goal was to focus instead on a specific cultural branch: Egypt. Starting anew, we began with only the thumbnail images used by the Oriental Museum for their AdLib system and the Discovery website. The thumbnail image data was easier to work with for a number of reason:

1) the 300x300 pixel size meant images could be more quickly viewed and transferred to class folders
2) the file system better structured ('archaeology', 'egyptian', 'oriental' as top level folders with subfolders organized by date and collection identifier)

3) the image data consisted only of objects from the museum collection, without rulers, color cards, or promotional materials

When creating a new data set for the current research, our specific focus was to try and group image collection into categories which contained several artefacts sharing homogeneous structure and stylistic motifs. For Egyptian anthropomorphic figurines we selected statuettes whose compositional stance was upright, with the left foot slightly forward. There was still some uncertainty in the labelling, in part due to Chimeric animal/human combinations within the Egyptian material, as well as some figurines with both animal and human as the subject matter. In those cases, we defaulted to the overall silhouette or material of the object to guide our final labelling choice.

After separating out the material into subtypes, we sought to revise and rebalance our image classes. New photos were taken of previously undocumented figurines: search terms such as 'figurine' and 'statuette' were used to search through the Discover webportal, along with some of the more common Object Typologies that had been observed in the existing image set. Those results which did not have a thumbnail image were noted, and the list sent to the Collections Registrar of the museum. An initial storage check was arranged, to examine the objects and see if they were suitable for photography (there was some initial concern that the reason they were un-photographed was due to extremely poor condition). Fortunately, many of the objects were in an intact state; further, their location within storage often resulted in finding additional previously un-photographed objects suitable for photography within the storage unit. Several documentation sessions were then arranged to take place within the museum's photo studio. This gave us access to the same lighting setup, background canvas, object stands, and the current DSLR camera being used by their photography team (Nikon Digital SLR D800, 36.3 MP AF, Sensor size: 861.6" CMOS); this helped to ensure parity with the existing collection training data.

One limiting factor that remained was in regard to the various types of Egyptian vessels (Lamps, Bowls, Canopic Jars, etc). While setting up several photo sessions for small, portable figurines was relatively easy to arrange with the museum, acquiring additional Egyptian pottery was not. This was due to several factors, including:

1) finding suitable (unbroken) vessels based solely on searching through the Discovery portal
2) the difficulty of retrieving such items from storage due to greater fragility
3) necessary oversight by museum staff during photography sessions (again, given the greater fragility)

The results of this work are shown in Table 1 and Figure 6.

Object Type	Total Images
Baboon	71
Bird	218
Bowl/Cup	97
Bull/Cow	71
Canopic Jar	120
Feline	143
Figure Head/Torso	116
Frog	23
Jug/Jar/Vase/Pot	490
Lamp	27
Scarab	260
Seated Egyptian Figure	173
Situla	26
Standing Egyptian Figure	313
Ushabti	323
	2471

Table 1 - Egyptian Subtype Image Data Set

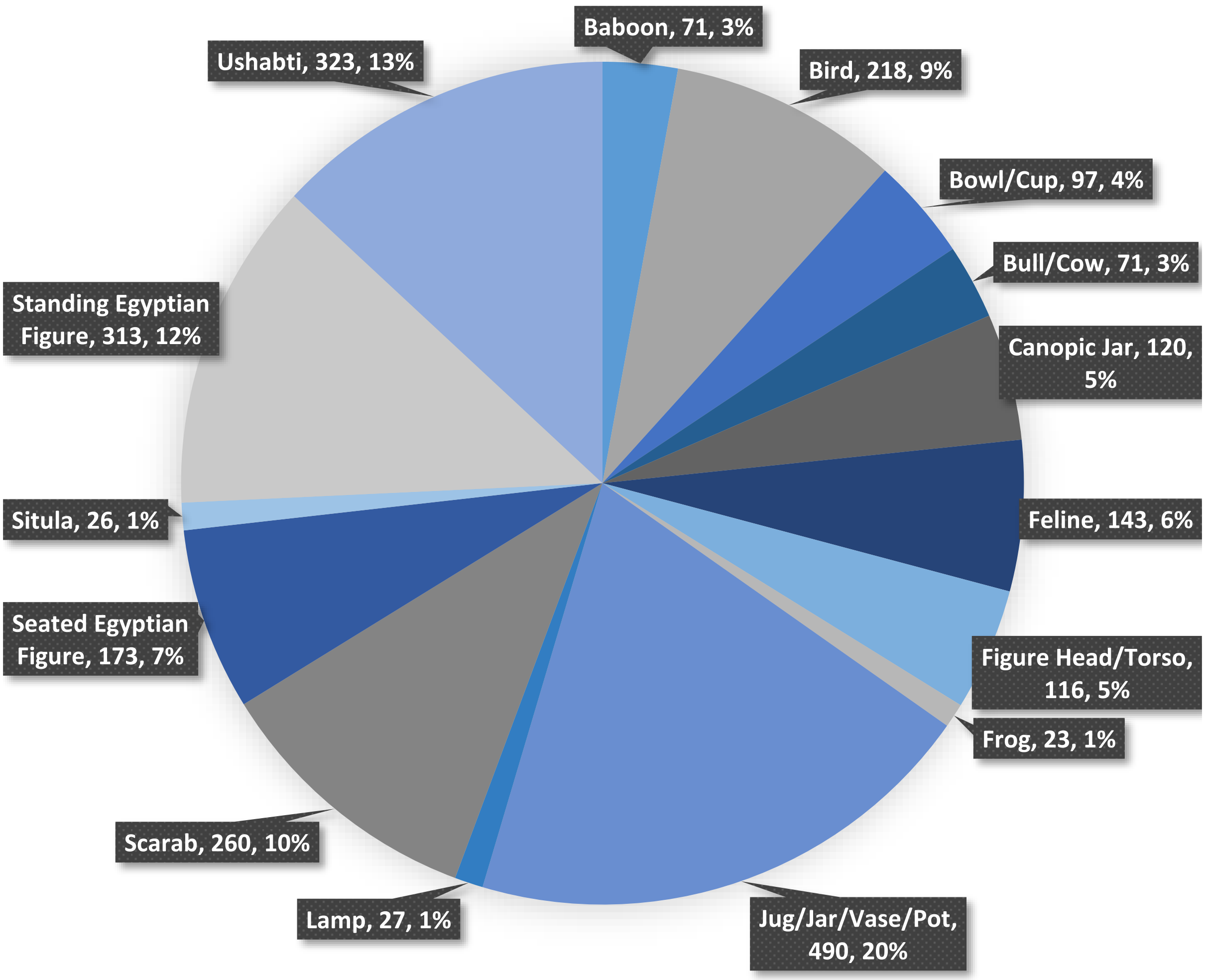

Figure 6 - Egyptian Artefact Type Image Data

Ultimately, we found that there simply weren't enough categories of images with similar enough structure and un-photographed material to be able to create classes of suitable size and distribution to be able to balance out our data. The previous summer research had shown class imbalance was a significant factor in the ability of machine learning software to produce accurate predictions through data training. Though the museum's Egyptology collection is extensive and fits within the general type of material we had hoped to use for our research, we had to revert to our previous distribution of objects into six categories, containing both a specific cultural branch (Egyptian) and a broadly generic branch (Asiatic).

3.2.3 Six Class Data Set Re-Development

Returning to the thumbnail images provided by the museum, we sought to apply the same sub-type breakdown to the material contained in the Oriental folder. We chose to include more material from India, including figurines seated in a lotus position, a common stance across several of the cultural statuettes in the collection. To try and maintain class balance, Asiatic Vessel imagery (of which previously there had actually been an overabundance) was reduced to roughly the same number of images. The animal figurine data was augmented with additional photography of small carved figures, primarily consisting of dogs (which became the majority sub-class in that grouping).

Our final resulting data set is shown in Table 2. It is based on broad collection provenance (Egyptian and Asiatic), followed by object type (anthropomorphic figurine, zoomorphic figurine, and containers or vessels). It consists of 5,000 labelled thumbnail images from the museum. As such, the six-class labelling was used as the baseline for training and testing, with the primary focus being on using it to develop a good CNN model structure which could be used for further experimental work.

	Total Number of Objects	Total Images
Asiatic Anthropomorphic Figurine	152	930
Asiatic Zoomorphic Figurine	96	817
Asiatic Vessel	109	765
Egyptian Anthropomorphic Figurine	176	925
Egyptian Zoomorphic Figurine	95	803
Egyptian Vessel	128	760
	756	**5000**

Table 2 - Six Class Image Data Set

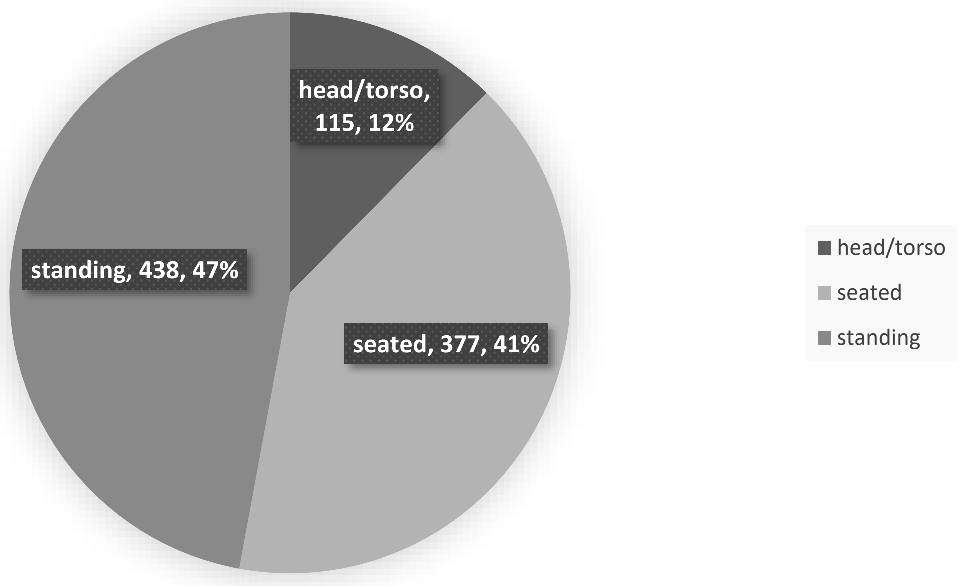

Figure 7 - Asiatic Anthropomorphic Figurine Class Composition

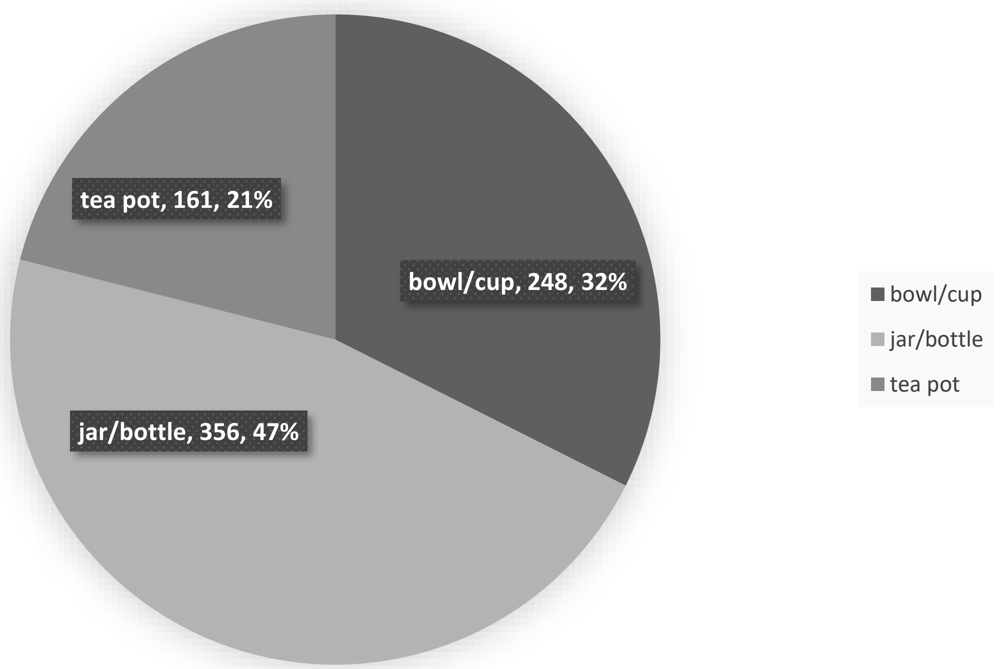

Figure 8 - Asiatic Vessel Class Composition

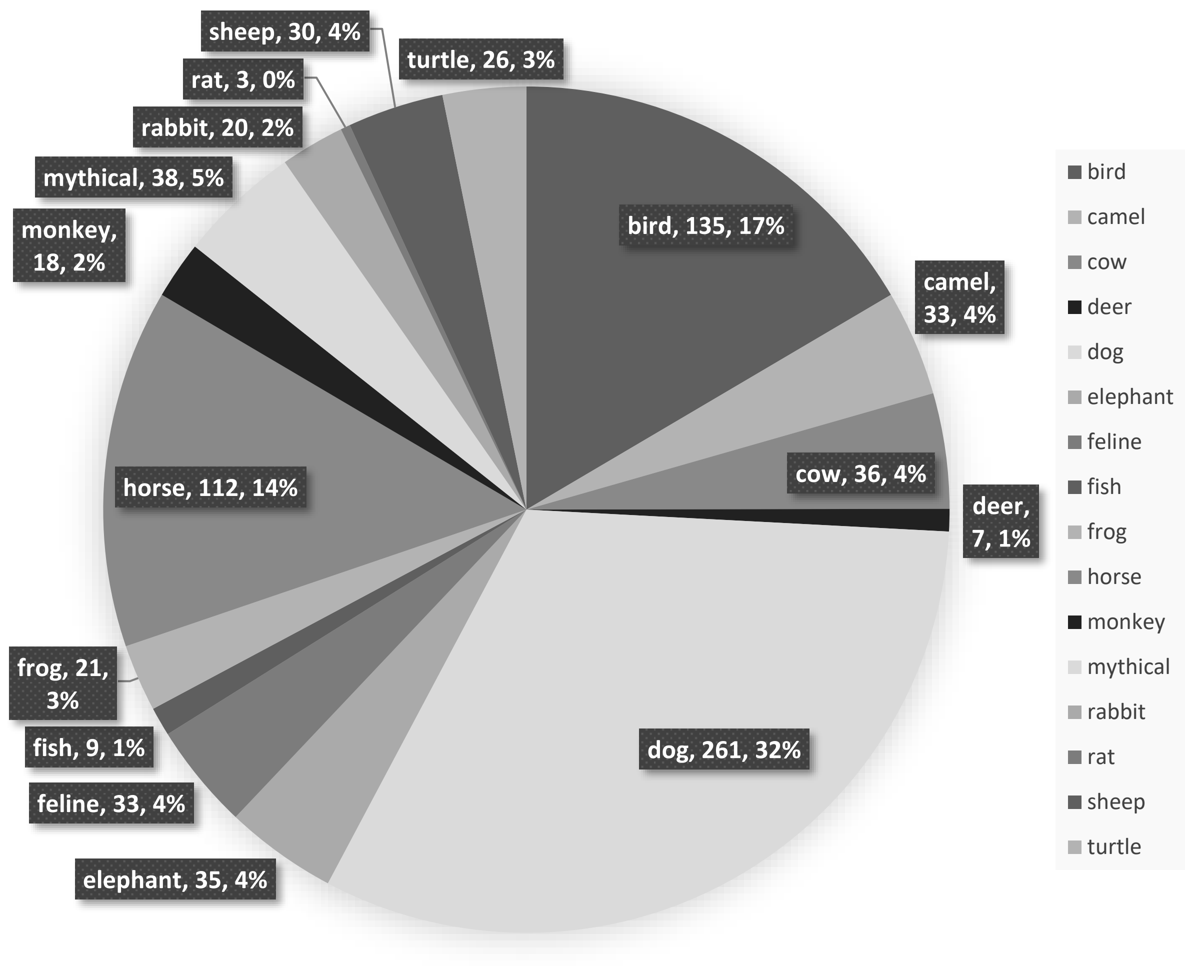

Figure 9 - Asiatic Zoomorphic Figurine Class Composition

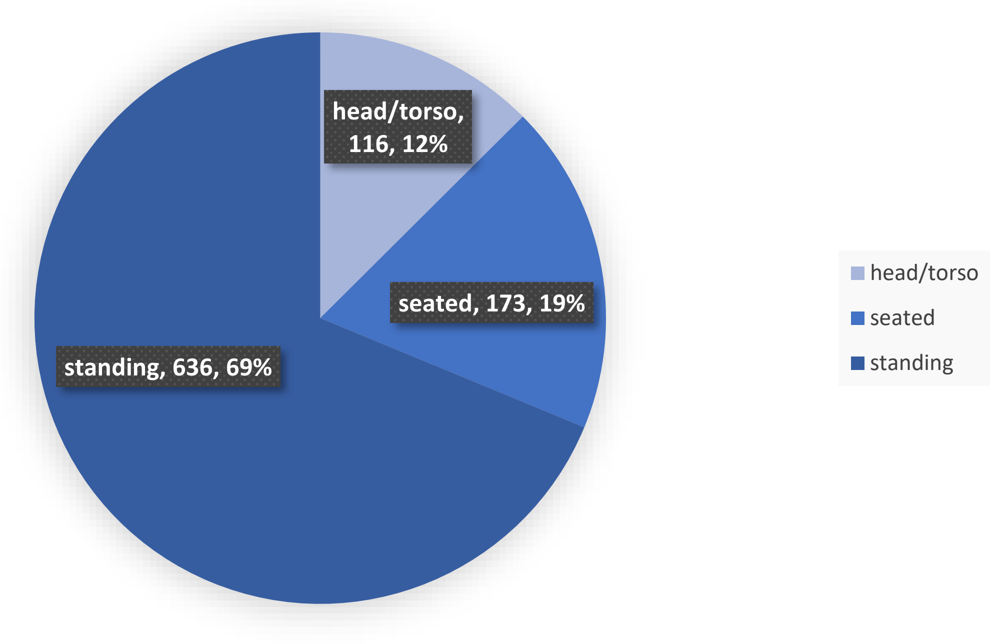

Figure 10 - Egyptian Anthropomorphic Figurine Class Composition

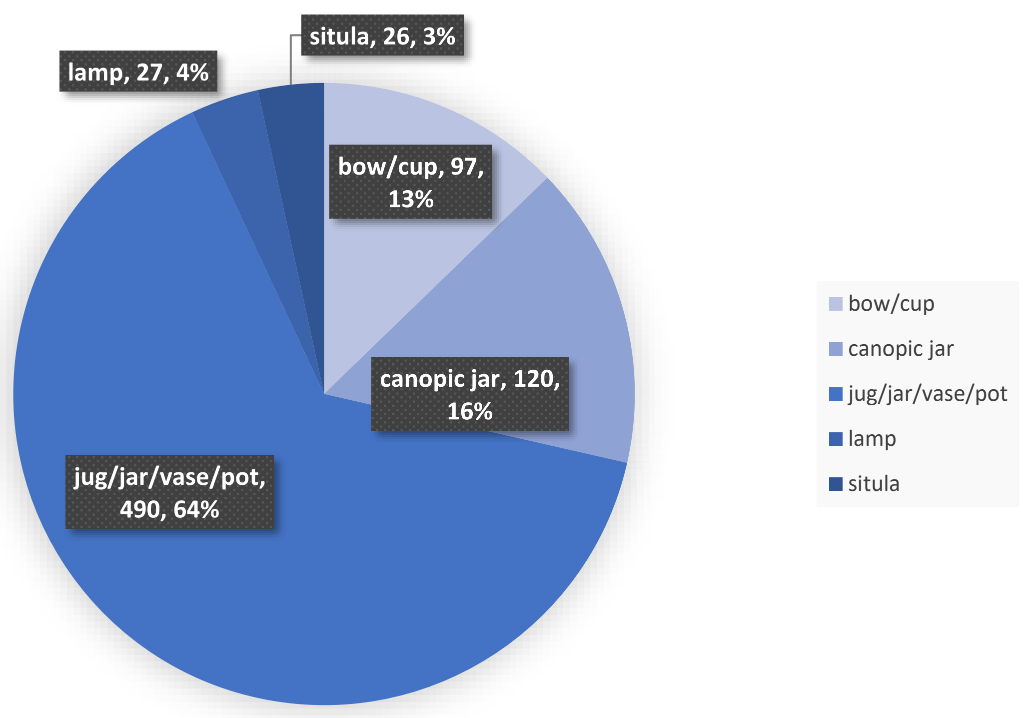

Figure 11 - Egyptian Vessel Class Composition

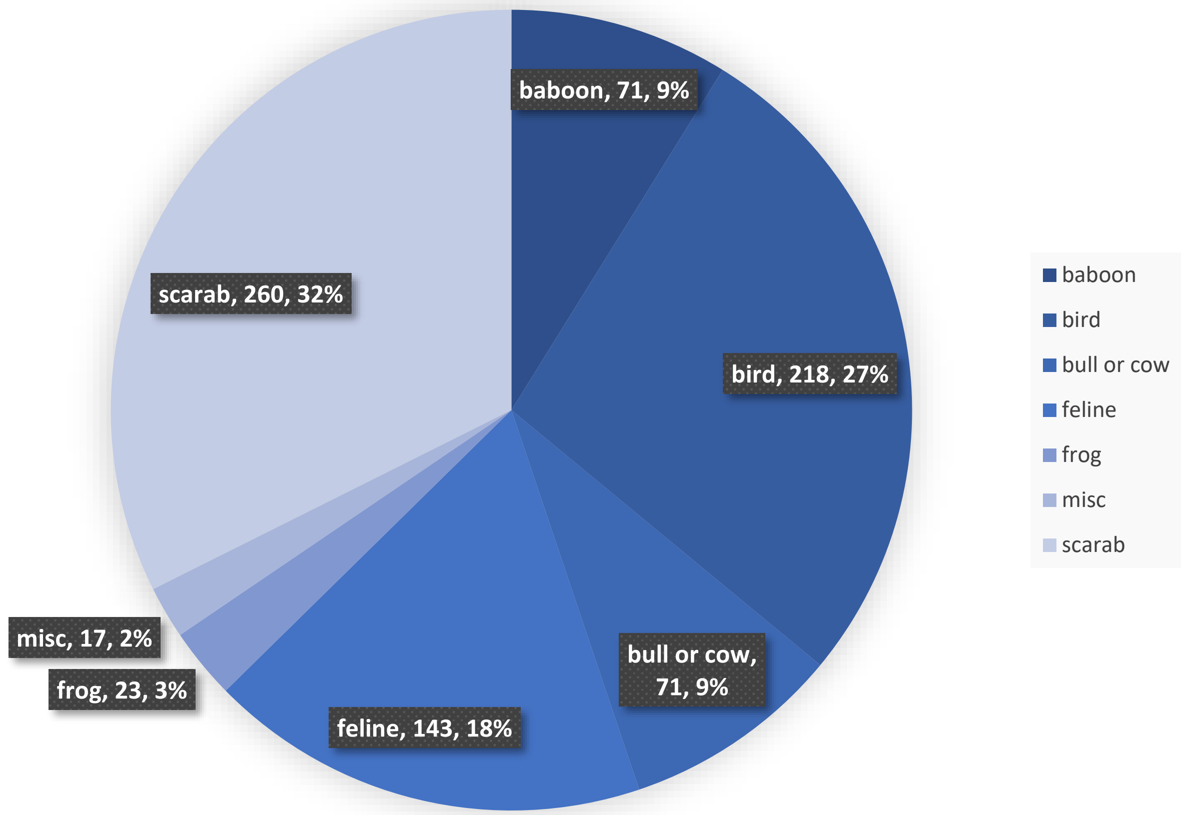

Figure 12 - Egyptian Zoomorphic Figurine Class Composition

3.2.4 Test Sets

Three test sets were created to check the performance of the network in various predictive tasks, based on three broad criteria groups:

1) Images of objects from the Oriental Museum of the same typology but not included in the training
2) Objects which were included in the training data thumbnails but photographed within display cases rather than ideal photolab conditions
3) Objects on display from other museums which are of the same broad typology, but which were not included in the training data (and subject to the same noisy data condition as the second test set above)

3.2.4.1 Test Set 1 – 500 Images

Taking the 5,000 labelled image data, we split out it into a training set of 4,500 images and a test set of 500 images. The test set was composed of typical object types and perspective images which demonstrated the style typology we hoped to optimize our classifier for. Image numbers for each class was kept at either 85 (figurines) or 80 (vessels), such that the total test images came to be 500 out of 4,500 total images.

	Training Images	Test Images	Total Images
Asiatic Anthropomorphic Figurine (AFA)	845	85	930
Asiatic Zoomorphic Figurine (AFZ)	732	85	817
Asiatic Vessel (AV)	685	80	765
Egyptian Anthropomorphic Figurine (EFA)	840	85	925
Egyptian Zoomorphic Figurine (EFZ)	718	85	803
Egyptian Vessel (EV)	680	80	760
	4500	**500**	**5000**

Table 3 - Classification Data Set (Training Set and Test Set 1)

3.2.4.2 Test Set 2 – Oriental Museum Display Cases

For the second test set, we collected new noisy data intended to check the real-world utility of our trained classifier. In order to do so we photographed objects located in the display cabinets of the Oriental Museum, and thus includes typical environmental variations:

- Variations in light intensity
- Glass reflections
- Camera angle
- Partial occlusions
- Background clutter
- Blurred photography

This type of noisy data was chosen specifically as a method of challenging the model, given that the overall desire for this research was to possibly deal in material identified in non-ideal situations (such as in the field or on display in an antiquities dealer shop).

We chose a similar number distribution for our classes as Test Set 1 (figurines balanced but with fewer vessel images), and collected the data using a Samsung Galaxy S8 camera phone (Dual Pixel 12MP AF, F1.7 Aperture, Sensor size: 1/2.55"). Some additional post-processing was necessary in order to prepare the resulting image files for the trained model. The photos were too large (averaging about 3 to 7MB, or 3024x4032 pixels), and had to be reduced in size. We were able to use the open-source image processing program Image Magick to batch process the image files, specifically making use of the “mogrify” command line function to resize all of the images to 300 pixels along the widest edge (resize=”300^>”), and to ensure the correct orientation of the converted photos using the automatic orientation flag (“-auto-orient”).

We chose objects representative of the dataset, though we were limited by what parts of the collection were currently on display. Given our research specifically sought to classify broad categories of objects, we feel that this limitation was minor in regards to our test focus. The primary difficulty was in matching the accession number to the correct object, as some of the display case tags did not include the identifier (this was mostly a problem for the Asiatic material).

	OM Display Case Images
Asiatic Anthropomorphic Figurine (AFA)	50
Asiatic Zoomorphic Figurine (AFZ)	50
Asiatic Vessel (AV)	50
Egyptian Anthropomorphic Figurine (EFA)	50
Egyptian Zoomorphic Figurine (EFZ)	50
Egyptian Vessel (EV)	50
Total	**300**

Table 4 - Test Set 2 Oriental Museum Display Case Images

3.2.4.3 Test Set 3 – United Kingdom Museums Display Cases

The final test set was created in the same manner as Test Set 2, however the focus here was to test the model on objects of similar type category from other museums in order to test how well the network generalizes to objects completely outside of the labelled training images. The museums chosen were the Bristol Museum, the British Museum Room 1 (the

Enlightenment Room), the British Museum Asian Gallery, the British Museum Egyptian Gallery, and the Victoria & Albert. The wide selection was in part necessary due to not all museums having appropriate types of objects (for instance, the V&A had a number of Asiatic figurine and vessels similar to those contained within the Oriental Museum, but no Egyptian objects). As with Test Set 2 the photos included environmental variations (lighting, occlusions, background clutter, etc), and were taken with the same camera phone. The same post-processing was necessary to prepare the data for the model. Unlike the Oriental Museum, a fully balanced set of classes was not possible (specifically, finding Zoomorphic Egyptian Figurines proved the most difficult). However, since the data is not intended to be used for training purposes (where class balance matters), it was decided that more was better, to provide as much data points as possible.

	Bristol Museum	British Museum Cultural Galleries	British Musem Room 1	V&A	Total
Asiatic Anthropomorphic Figurine (AFA)	22	46	54	34	**156**
Asiatic Zoomorphic Figurine (AFZ)	9	44	0	49	**102**
Asiatic Vessel (AV)	9	53	34	29	**125**
Egyptian Anthropomorphic Figurine (EFA)	43	122	87	0	**252**
Egyptian Zoomorphic Figurine (EFZ)	6	25	8	0	**39**
Egyptian Vessel (EV)	13	70	16	0	**99**
	102	**360**	**199**	**112**	**773**

Table 5 - Test Set 3 United Kingdom Display Case Images

3.3 Evaluation Criteria

3.3.1 Prediction Metrics

In order to judge the performance of the models on our various test sets, we chose the statistical metrics for informational retrieval most commonly employed in machine learning (Sokolova, Japkowicz and Szpakowicz 2006). Though the use of these performance metrics are not without criticism (for example in the referenced paper by Sokolova, Japkowicz, and Szpakowicz), the fact that our classes are intended to be evenly spread with no particular class of interest supports the use of the Precision and Recall metrics as appropriate evaluation criteria.

		Predicted Class	
		Class = Yes	Class = No
True Label Class	Class = Yes	**True Positive (TP)**	**False Negative (FN)**
	Class = No	**False Positive (FP)**	**True Negative (TN)**

Precision: Number of True Positives divided by number of True Positives + False Positives. Number of correct predictions divided by the total number of predictions for the class.

$$P = \frac{TP}{TP+FP} \tag{1}$$

Recall: Number of True Positives divided by the number of False Negatives. Ratio of correctly predicted classes to all observations of actual class.

$$R = \frac{TP}{TP+FN} \tag{2}$$

F-Score: Weighted average of Precision and Recall.

$$F1 = \frac{\left(2*(R*P)\right)}{(R+P)} \tag{3}$$

3.3.2 Confusion Matrix

This is a popular method used to display the True Labels and Predicted Labels on a matrixed grid for the test set, with coloration to show concentrations of the data. Sometimes it is called an Error Matrix; its purpose is to show where a machine learning prediction algorithm may be confusing actual and predicted classes. Like the metrics above, it is a common method for interpreting the results of a classifier. The ideal situation is a solid diagonal line across the top left down to the bottom right, representing a high number of True Positive results from the model when the test data is run through it. Normalization of the Confusion matrix is useful when there is an unbalanced class test data, showing the percentage of classifications rather than the raw prediction data.

3.4 CONCLUSION

The Oriental Museum's digital records reflect the content and guiding focus of that institution, and of its history. A clear understanding of where it came from provides a better context for the records contained therein, how certain artefacts and objects might be over and under-represented in the museum's digital records. Our first research question asks how well a curated, pre-existing museum collection image data set serves for typology classification. The first step in answering that is whether we can create an image data set serviceable for the current state of machine learning technology. Our initial hope to focus on a single cultural branch and clear source of materials (Egyptian antiquities primarily provided by the Northumberland and Wellcome collections) was unsuccessful, as diving into the digital image data did not produce enough representative data points to begin work. Instead we had to rely on a similar classification type structure as was conducted in previous work, revised and expanded. A new Classification data set was created, with a total number of 5,000 digital images from the Oriental Museum organized into a specific cultural branch (Egyptian) and a broad cultural branch to contrast (Asiatic). Classes were based on broad types of objects inspired in part by the organizational divisions of ObjectID and the ICOM Red Lists (human and animal figurines, vessels), with sub-categories more precisely describing the artefacts contained within. Proceeding from there, three test data sets were created: a base case (500 images split out from the 5,000 image data set), and two challenging cases. The challenging cases consisted of new images recorded from museum display cases, developed based on the type of real-world scenarios which could challenge an image recognition system (namely objects in display cases with multiple variations of lighting, angle, reflections, and occlusions). Further, the test cases were split into one set of objects that were in the training data set (Oriental Museum display case images) and similar classes of objects that were not part of the training (display case images from multiple other museums). Finally, a set of evaluation criteria was chosen based on common methods for judging image retrieval in data science, including prediction metrics (Precision, Recall, and F-Score) and visual display (Confusion Matrix). Using these criteria provides a quantitative set of measures to judge the performance of typology Classification predictions. From here, it is possible to proceed to the next step of evaluating our first research question, to begin experimenting with machine learning computer vision.

4 AUTOMATED RECOGNITION: OBJECT CLASSIFICATION

4.1 SUPERVISED MACHINE LEARNING

Supervised Learning is a branch of **Machine Learning** (ML) in which labelled data is used to approximate a function to map training data to a set of defined targets. The resulting inferred mapping then is used to predict the labels for new, unseen data samples (of a similar type to the original training data). This is different from the traditional programming paradigm, wherein the programmer seeks to manually craft a set of rules and algorithms that can generate a desired output from the input data. This is useful for image processing, as the process of trying to create a sufficiently flexible set of rules to recognize all potential variations of an object within a digital photograph in an attempt to mimic the human capability of visual recognition quickly becomes incredibly complex to develop by hand (this is the *semantic gap* referred to in Section 1.1). Instead, for ML the goal is to use training examples of the desired input-output behaviour to let the computer optimize a set of mapping rules according to a define performance metric samples (Jordan and Mitchell 2015). Practical software applications can thus be created which are more robust to variations of input data, as the ML algorithm optimises network parameters to generate an output as accurately as possible based on that input.

Supervised Learning systems implement the above concept by taking in training data in the form of (x, y) pairs, with the goal of producing a prediction y in response to a query x, using a learned mapping $f(x)$. The (x, y) pair in practice forms the **training data**, consisting of training samples matched with corresponding outputs to be utilized during the training process. Like this is the **test data**, a matching set of samples and outputs in the same domain but which are specifically not used during the training phase, an unseen dataset that is used to test the performance of the learned mapping. Sitting between these two is the **validation data**, similar in many respects to the test data in that it is used to evaluate the learned mapping but specifically utilized to evaluate the performance of the learning function during the training phase (in order to spot if the parameter weights are fitting above or below the target prediction). Finally, the $f(x)$ learned mapping or function (also called the **inferred function** or the **model**) is the probability distribution which produces output y given input x.

Systems capable of implementing the above concept are varied, with **Convolutional Neural Network** (CNN) models being a particularly high-impact focus of current research (specifically in relation to computer vision and image processing, as discussed in the Introduction). Currently, this type of machine learning is considered an extremely promising method for the automatic extraction of meaningful information from image data, with broad cross-disciplinary potential (LeCun, Bengio and Hinton 2015) (Goodfellow, Bengio and Courville 2016).

4.2 Convolutional Neural Networks

As mentioned in the previous chapter, while the wide-scale popularity of CNN's has occurred over the past ten years, they have been used for computer vision recognition systems since the late 90's (specifically, Yan LeCun's landmark experiments at Bell Labs using a backpropagation-algorithm and gradient-based learning CNN for automatic character and digit recognition) (LeCun, Bottou, et al. 1998). The structure for these initial CNN's tended to be simple, consisting of a series of stacked layers performing image classification by initially looking at low-level features (such as edges and curves) and then building up to more abstract pattern matching.

The name Convolutional comes from the algorithmic function by which each of these layers produces an output, passing a small matrix (or **kernel**) containing numerical parameters (a combination of mathematical **weights** and **biases**) over the 2-Dimensional matrix of pixel values which comprise a digital image. As the kernel is applied to each region of the 2D input data, it performs elementwise multiplication that is then summed up into a single output value.

As an example, consider Figure 13. Here we have a simple 5x5 black and white image, in which 0 represents a white pixel and 1 a black pixel. The kernel is a 3x3 matrix which slides over each section of the image, multiplying the pixel values by the kernel values, resulting in a single number output that represents the values contained within that section of the image. The goal of this is to filter an image, looking for patterns represented by the output feature matrix.

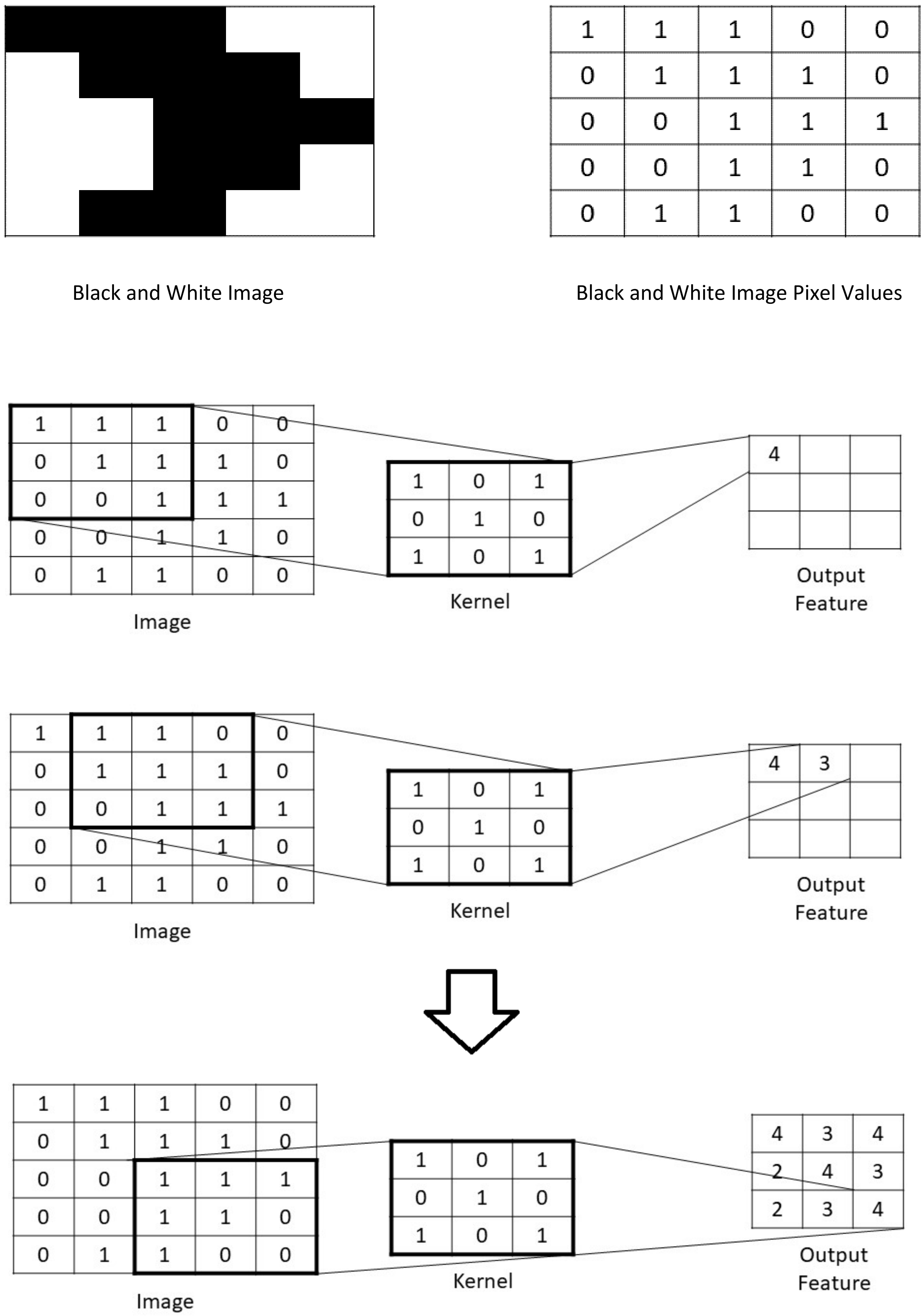

Figure 13 - Example of a 3x3 Convolutional Filter being passed over a 5x5 black and white image, resulting in a new 3x3 matrix of output feature values

This process of passing a filter over an image is called **convolving**, and is where the convolutional in Convolutional Neural Network comes from. The output of the convolution will be smaller (both in width and in height) from the original image, and in CNN's there will be a linear function applied between the kernel and the image window it is laid over. These are called **Convolutional Layers**. Similar to this are **Activation Layers**, where the output is called an **activation** (inspired in part by how a biological neuron fires based on visual stimuli, and thus providing the Neural for a Neural Network). The mathematical formulation for this **Activation Function** is:

$$f(x, W) = Wx + b \tag{4}$$

Treated as a **feature vector**, here *x* represents the input data from the filter, *W* is the weights by which that data is multiplied, and *b* is some independent value which serves the purpose of influencing the results towards a desired output. It is important to note that (unlike a Convolutional Layer), the Activation Layer must introduce non-linearity to the network's internal layers. This is because a linear activation does not benefit from the stacking of multiple layers, and ultimately provides results no different from a single layer perceptron network (in other words, if all of the activations of a network are linear, then the final activation is equivalent to nothing more than a linear activation of the input of the first layer).

During the training phase of a CNN network, these weights and biases are applied according to some algorithmic set of methods to develop a system of activations which maximizes the prediction accuracy of image inputs. In essence, the machine learning algorithm is attempting to find a set of parameterized activations such that when an image passes through all layers of the network, it will maximize the number of correctly mapped training inputs to their associated labels (the **Optimizer**) by measuring some metric to calculate how far off a prediction is from the true target (the **Loss Function**). Each filter can be thought of as a feature identifier, the method by which image information expressed by pixel values can be recognized; the purpose of the filter is to output a high number to a multi-dimensional array (called a **feature map**) in regions of the image which contain the identifying feature. This output is in turn passed along to the next layer, such that as the data progresses through the layers you will get increasingly more meaningful transformations of the data, thus learning more useful representations of the input image (and thus identifying more complex features). Initially, the weights for each layer are given a random value; with every pass or iteration through the network layers (called an **Epoch**), the weights are adjusted to attempt to achieve a better prediction score by minimizing the loss and updating the weights through the optimizer. The ideal is a network producing outputs that perfectly match the label targets for all input images; the process by which the network seeks to achieve that ideal is through the Optimizer and Loss Function (see Figure 14).

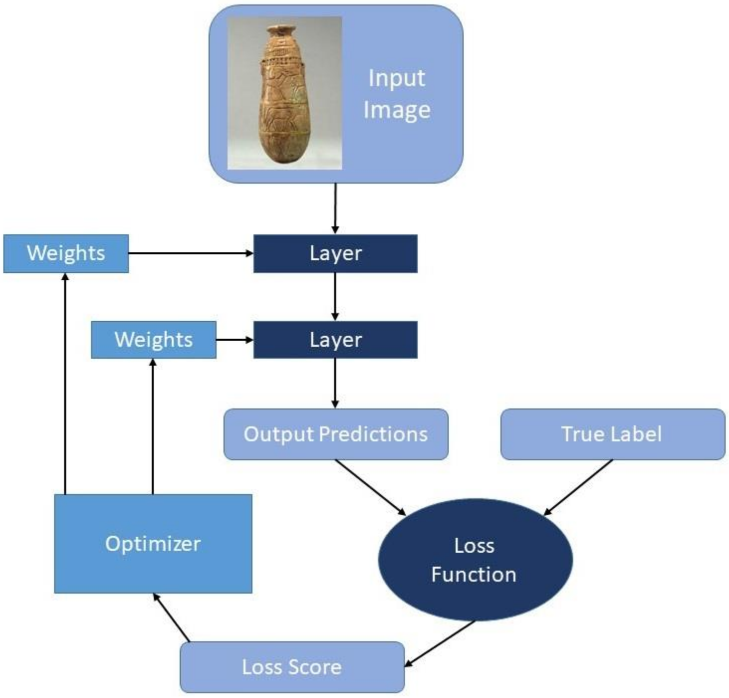

Figure 14 – Backpropagation Conceptual Data Flow, source Author

Shallow CNN's have only have three layers: an Input Layer, a Hidden Layer, and an Output Layer. The **Input Layer** receives the digital image, the **Output Layer** produces a single class or probability of classes that best describes that image, and the **Hidden Layer** is simply an internal layer that is neither an Input or Output Layer. The term Deep Learning comes from the more recent development of CNN architectures which stacks increasing numbers of layers on top of one another (Deep vs Shallow). A modern Deep Learning CNN takes in an image as an input and then passes it through internal hidden layers which are first identifying low-level features, then mid-level features, and finally high-level features. Other operations are also performed by these hidden layers, such a Rectified Linear Unit processing (**ReLU Layer**), dimensionality reduction (**Pooling Layer**), and randomized neuron connections to reduce overfitting (**Dropout Layer**).

A ReLU layer is a specific type of Activation Function which implements a maximizing function to remove negative numbers from the output of previous layers and replaces them with zero, as shown in Figure 15:

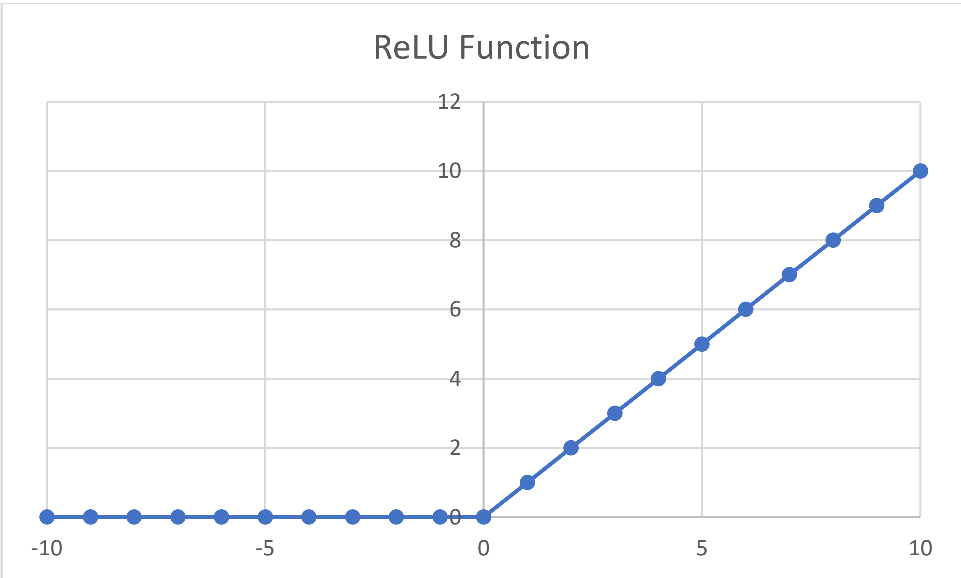

Figure 15 - Maximizing Effect of Rectified Linear Unit on input data

This can be expressed as the equation

$$Relu(z) = \max(z, 0) \tag{5}$$

with the max function further defined as:

$$max(z) = \begin{cases} z, & z > 0 \\ 0, & z \leq 0 \end{cases} \tag{6}$$

There are similar functions which can be used to reduce the previous layer's output. Two common examples include the Sigmoid function, which takes in the previous layers values and outputs a value in the range of [0, 1]:

$$Sigmoid(z) = \frac{1}{1+e^{-z}} \tag{7}$$

as well as the Tanh function, which produces output values between the range [-1, 1]:

$$Tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \tag{8}$$

Many modern neural networks default to the ReLU activation; the Sigmoid and Tanh functions are susceptible to a problem known as a vanishing gradient, in which the gradient (the multi-variable generalization of the derivative for the function) returns values from the loss function which approach zero as more layers are added (which increases the difficulty of training the network). The ReLU function avoids this problem, as well as being less computationally complex due to simpler mathematical operations. (He, Zhang, et al. 2015) (Goodfellow, Bengio and Courville 2016, 194, 226) (Rawat and Wang 2017).

In addition to the activation layer, the Pooling and Dropout layers aid in the training process. Pooling Layers are used to reduce the complexity of the feature maps produced through convolution and ReLU processing, in order to decrease the memory space of the millions of results produced by a neural network training period. Common Pooling strategies are **Max Pooling** (in which the largest value of a kernel window is selected as the output) and **Average Pooling** (in which the average of all values contained within the kernel window is used as the output). Dropout is a technique that improves network performance (Srivastava, et al. 2014) by randomly deactivating nodes within the network's forward and backpropagation training cycle. This increases the number of epochs necessary for training the network, but in exchange forces it to calculate stronger feature identification weights and biases.

At the top layer of the neural network, the feature map output passes through a layer which should reduce the high-dimensional data down to a predicted classification label for the image. This final layer is fully connected to all of the processing layer outputs before it (and is thus called a **Fully Connected Layer**), and produces an N-dimensional vector of N classes (N being the number of class labels) with each number in the vector representing the probability of a certain class. These probabilities contained within the final vector may be real number values which are negative or greater than 1. Often a **Softmax** function is then applied to calculate the probability distribution of the predicted output classes such that each component of the vector is then bounded within a (0,1) interval, and scaled such that the components sum to 1. The method by which the Softmax function produces this for a given probability contained within that vector is shown below:

$$\sigma(z_j) = \frac{e^{z_j}}{\sum_i e^{z_i}} \tag{9}$$

To summarize, Convolutional Layers are used to multiply an image by a kernel value of weights (optimized using gradient descent), Pooling and Dropout Layers are used to reduce the dimensionality of the layers they are applied to, and Activation Layers are used to squash the values of a layer into some range, such as [0,1] or [-1,1] (Reppel 2017). In practice we can see these implemented in one of the earliest and most straightforward convolutional networks: LeNet-5 (LeCun, Bottou, et al. 1998), shown in Figure 16. Used for text recognition, it demonstrates a simple network using convolutional and average pooling layers to generate stacked feature maps of reduced dimensionality, followed by a fully connected layer and a final softmax classifier for output predictions. Modern CNN Models use more complex architectures, but at the core these are the building blocks which compose any Convolutional Neural Network.

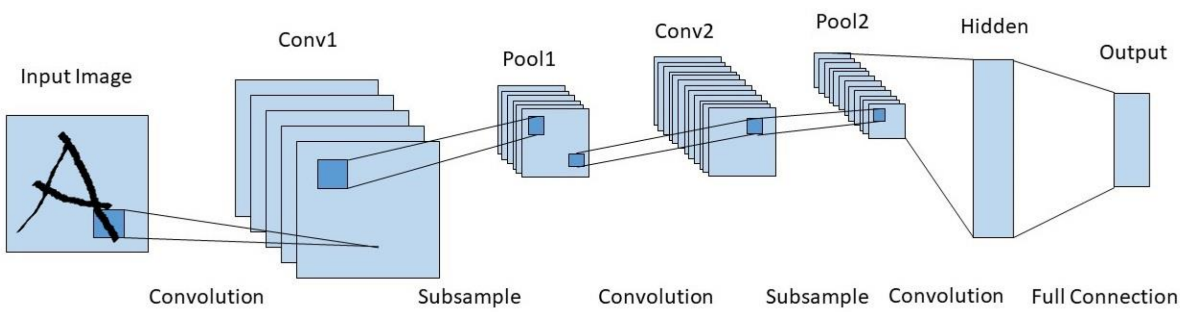

Figure 16 – LeNet-5 Convolutional Neural Network Architecture, proposed for Optical Character Recognition in the 1990's, based on source from the original paper (LeCun, Bottou, et al. 1998)

4.3 METHOD

4.3.1 Transfer Learning

Given that our training and primary test image data combined only comes to 5,000 individual images, we are slightly restricted in the method by which we can train a neural network. Fortunately, there is a common and demonstratively effective approach for creating a CNN classifier when the training data is considered small (i.e. less than 10,000 images) that has found wide-spread usage within the machine learning and computer vision community: **Transfer Learning** (Pan and Yang 2009) (Taylor and Stone 2009) (Shao, Zhu and Li 2015) (Weiss, Khoshgoftaar and Wang 2016). Unlike in traditional neural network training wherein the weights and biases of the network are initially set to randomized values, a pre-trained CNN begins with its parameters set from those previously developed by lengthy training on a very large image dataset. This method works best when the source domain is similar to the new target domain (in our case, image data that has similar structures). Conceptually, this method is primarily effective due to the layered nature of a CNN acting as feature extractors of increasing complexity: different layers of the model are trained to activate based on detected image representations (blobs, curves, lines, colors, textures, etc). These same image representations can be repurposed to the new target domain, without needing to be retrained for. These generic detector layers form the **convolutional base** of the new CNN model, with the weights and biases frozen during the training period. Instead a new set of layers are appended to the end of the existing network, and trained to act as the new classifier. Training of the CNN Model then follows the same backpropagation strategy described above.

In this type of Supervised Learning, there is a three stage process for evaluating CNN models. The image data set is split into Training, Validation, and Test Sets (as described previously). The CNN is trained on the Training Data, and then evaluated using the Validation data that has been set aside specifically to keep the network from learning the Training Data too well, as this can lead to an inability to generalize to data it has never seen before (a phenomenon called **Overfitting**). The Training period switches back and forth between the two data sets: training on the Training data, evaluating using the Validation data, until finally a model is produced with a high **Validation Accuracy**. Such a model is saved, and can then be loaded and run over one or more test sets to check how well the model's predictive capability actually is (this application of the model to never-before-seen data is called **Inference**).

4.3.2 CNN Model

Our image classification CNN model was implemented in PyTorch, an open-source machine learning library for Python (Paszke, et al. 2017). One of its core strengths is an easily loaded set of Neural Network models pre-trained on the ImageNet Large Scale Visual Recognition

Challenge Dataset, containing 1.4 million labelled images mapped to 1000 classes (mostly animals and everyday objects). Given that our own data set is specifically focused on objects containing figurines with human and animal features along with household vessels, the source and target domains should serve as a strong baseline for transfer learning. By including the torchvision.models subpackage, we can rapidly download any of the currently available CNN architectures (AlexNet, VGG, ResNet, SqueezeNet, DenseNet, Inception v3, GoogLeNet, ShuffleNet v2) with the pre-trained weights cached in a local directory.

4.3.2.1 VGGNet

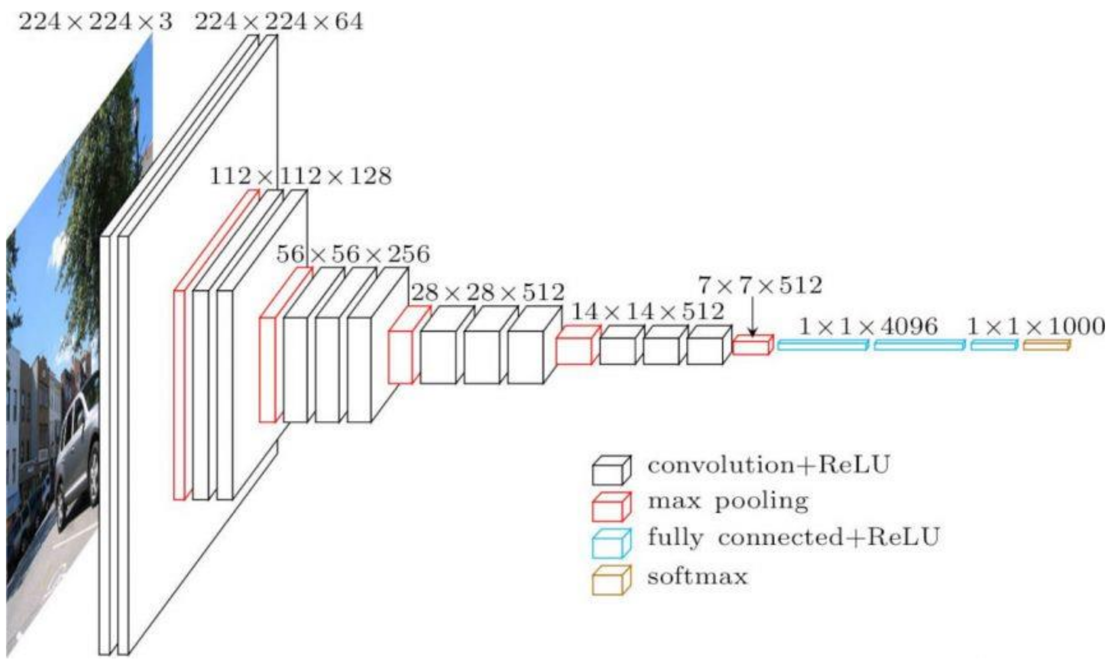

Figure 17 - VGGNet16 Architecture, source Davi Frossard (Frossard 2016)

This neural network architecture was developed by the Visual Geometry Group from Oxford University, in 2014. Inspired by the network architecture developed by Khrezhevsky and the Supervision group in 2012 (AlexNet), they improved upon the basic neural net design by replacing AlexNet's larger kernel filters (of sizes 11x11 and 5x5) with smaller stacks of 3x3 filters, showing marked improvement and winning that year's ImageNet challenge (Simonyan and Zisserman 2014). Several variations of VGG are available through the Torchvision model, however VGG16 and VGG19 are the two which are actually used for visual recognition (the smaller networks instead being used to incrementally train network configurations of increasing depth, as described in Simonyan and Zisserman's paper). We chose VGG as one of the models we would test due to it being a good example of a fairly straightforward CNN architecture (when compared to more recent models).

4.3.2.2 ResNet

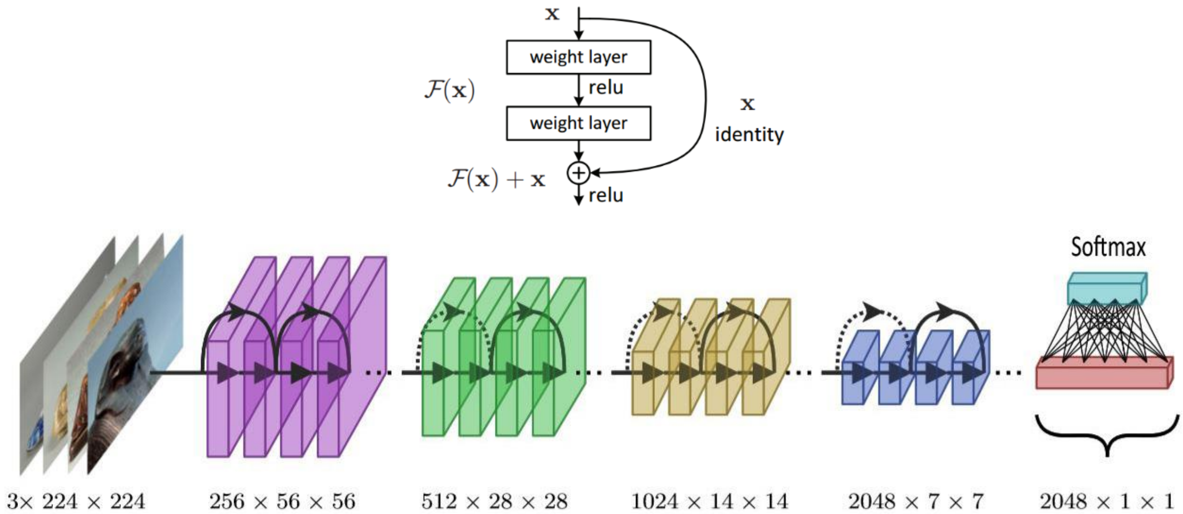

Figure 18 - ResNet skip connection building block, source original paper (He, Zhang, et al. 2016) and ResNet50 Architecture, source Author and Dr. Chris Willcocks

For our architectures we primarily chose to focus on variations of a deep residual network (ResNet), a CNN architecture first demonstrated in 2015 (He, Zhang, et al. 2016). ResNet has three levels of structure: the smallest building block is a single convolutional layer, batch normalisation and a ReLU activation. The next level is a residual block, comprised of two of the aforementioned modules with a residual skip connection over them. A skip connection simply copies the input to a function and adds it to the output. This element-wise addition is a way of passing state information from one ResNet building block on to other building blocks further down the forward propagation path (specifically, it is believed to be a way of helping to promote gradient forward propagation, as described in the Optimization section below). Several residual blocks are stacked and some of the convolutions are given a stride (a metric for describing the movement of the kernel for pixel-wise operations across a given image) of 2; this serves the same purpose as max pooling with conventional architectures, reducing the spatial dimension and increasing the feature dimension. We chose ResNet as our primary architecture due to current research suggesting residual connections provide a significant improvement in training a neural network (Rawat and Wang 2017). The available models of it are named according to increasing depth of layers: ResNet18, ResNet32, ResNet50, ResNet101, and ResNet152, all of which were incorporated into our testing.

4.3.2.3 DenseNet

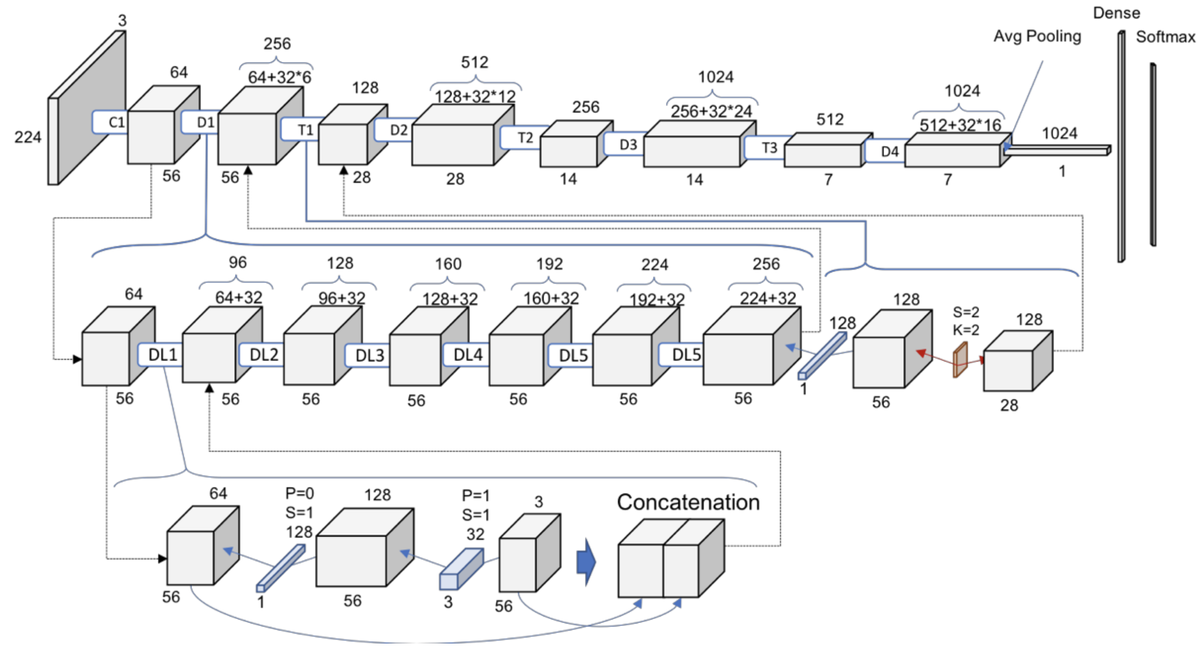

Figure 19 - DenseNet121 Architecture, source Pablo Ruiz (Ruiz 2018)

DenseNet is one of the most recently released CNN model architectures (Huang, et al. 2017), which has been released and made available in the time since our initial experiments with ResNet in the summer of 2017. Like with ResNet, the architecture passes information from previous module blocks forward through the neural network architecture. However, whereas ResNet layers pass that information forward utilizing skip connections from one layer to the next, in DenseNet each layer has connections to all the preceding layers. Whereas ResNet uses elemental addition, DenseNet instead uses concatenation to incorporate this data into downstream layers. This reduces the complexity of the architecture by increasing the breadth of the layers, simply creating a new set of feature maps with the previous feature maps stacked alongside. As with ResNet, this improves the gradient flow of the network, but also creates a network which detects more diversified features and increasingly complex patterns (Huang, et al. 2017). While our primary testing was done using the variants of VGGNet and ResNet, we also experimented with the version of DenseNet with a high performance score.

4.3.2.4 Final Sequential Layer

The final fully-connected layer appended to the end of the CNN model. The weights for the lower layers remain frozen, while this final layer is where the training takes place. The structure of this final layer is primarily the same for all the models we tested (VGG, ResNet, and DenseNet):

- Linear Layer - The input for the first linear layer is dependent on the final CNN convolutional base feature map output:
  - **VGGNet16, 19**: (input 25088, output 512)
  - **ResNet18, 34**: (input 512, output 512)
  - **ResNet50, 101, 152**: (input 2048, output 512)
  - **DenseNet161**: (input 2208, output 512)
- ReLU Layer
- Droput Layer (Probability of dropout set to 20%)
- Linear Layer (input 512, output Number of Classes)
- Log Softmax

4.3.3 Image Data Input

4.3.3.1 Data Loader Automated Training/Validation Split

The requirements for PyTorch image inputs is that all image files be 3-channel RGB images with Height and Width of at least 224 pixels (and of equal length), and normalized using a mean of [0.485, 0.456, 0.406] and standard deviation of [0.229, 0.224, 0.225]. The Data Loader utility from the PyTorch was used to retrieve images from the training data set folders. PyTorch uses the folder names to define the class labels for images loaded in this way. The thumbnail image size was ideal for processing, as images could be rapidly re-sized by the dataloader to the 224x224 pixel dimensions for the CNN Input Layer (larger image sizes can be used, but increase the processing time due to PyTorch needing to do more re-sizing). Another capability provided by the Data Loader utility was to randomly select 20% of the data set for use as the validation set.

4.3.3.2 Image Transformations

Images were kept in their original orientation, but were resized to a square aspect ratio by stretching the narrow edge. This was done instead of a central cropping in order to preserve all pixel image data, rather than risking losing edge parts of the object image. It also was used in preference to padding the narrow edge of the image out, as we were concerned that this might create a false training feature based on orientation rather than features inherent to the object within the input image. The mean/standard deviation normalization described above was applied at this time as well.

Since the training data is considered small (for this sort of supervised learning function), we applied a set of random image transformations to the data in order to artificially inflate the number of different images being used for training (a technique used to prevent overfitting and to improve generalization). The random set of transformations to be applied included:

- Flipping the image along the horizontal and/or vertical axis
- Random rotation of the image up to 180 degrees
- Randomly converting an image to grayscale
- Color change of the image brightness

These transformations were chosen based in part based on the sort of noisy data we hope our model will be able to account for (such as dimly lit display cabinets or a black-and-white photo).

4.3.4 Backpropagation Architecture

4.3.4.1 Loss Function

As previously described, the training loss function is the method by which our network evaluates how well it is modelling predicted data to an expected target (the classification label). It is an algorithmic method by which we can quantify the prediction scores being produced by the weights and biases of the network, to then be able to judge and act upon those results. When the resulting quantification output is large (indicating a poor prediction), the loss function will then work with the network's optimizer algorithm to modify the network weights and biases to gradually improve the output (i.e. minimizing the loss).

For classification problems the **Cross-Entropy Loss** function is used, as other loss functions are primarily geared towards regression (i.e. the prediction of continuous values such as word completion for search terms) (Goodfellow, Bengio and Courville 2016). The benefit of

Cross-Entropy Loss is that it produces quantification results that trend towards zero as the network results get better at predicting the labels, while increasing exponentially as the predicted probability diverges from the actual label. Given p as the set of actual labels anq q as the set of predicted labels for a set of measures X, the loss can be calculated as:

$$H(p,q) = -\sum_{x \in X} p(x) \log q(x) \qquad (10)$$

In PyTorch, the Cross-Entropy Loss function can be implemented using the built-in method for negative log likelihood loss (torch.nn.NLLLoss) combined with a Log Softmax layer appended to the end of the network (as was done in our network's final sequential layer, shown below).

4.3.4.2 Optimizer

While multi-class loss functions have a standard implementation, optimization functions are more varied within current CNN architectures. However, the most popular and commonly used optimization strategies for use with for neural networks are variations of Gradient Descent (Ruder 2016). Gradients are partial derivatives measuring change in the slope of a function, with **Gradient Descent** a method of calculating what direction the function's parameters must be changed based on the derived slope of the function in order to find its minimal value. One of the most commonly used analogies is that of walking across a valley, trying to find the lowest point by seeking out downward sloping paths. PyTorch (and all other state-of-the-art deep learning frameworks) provide automatic differentiation functions to calculate the gradient. For our implementation, we considered two of the most common Gradient Optimizations strategies: Stochastic Gradient Descent and Adaptive Moment Estimation.

Stochastic Gradient Descent (SGD) – For standard Gradient Descent (also called Batch Gradient Descent) the gradient of the loss function is calculated in regards to all parameters used in a forward pass through the network. Given the size of the feature maps involved, this strategy is generally avoided. Instead, SGD is a particular implementation of gradient descent which reduces the processing needs of the backpropagation optimization; instead of computing the gradient of the loss function by summing up all of the loss outputs, it randomly (stochastically) uses a subset of the training data for calculating the gradient. This reduces the computational overhead, while still demonstrating a capability of converging to the minima of a functions spatial surface. Krizhevsky's implementation of AlexNet used a variant of SGD incorporating the concept of momentum, a parameter added to the Optimizer update based on previous gradient calculations (Krizhevsky, Sutskever and Hinton 2012).

Adaptive Moment Estimation (Adam) – One variant of SGD is the Adam function. It also uses past gradient calculations to implement the concept of momentum, utilizing a moving average of the gradient rather than the gradient itself. It combines this with an exponentially decaying average of past squared gradients calculated across all parameters (an implementation common to several other optimizers developed at the same time, including Adagrad and RMSProp) (Kingma and Ba 2014).

In practice, Adam has been shown to improve training times and convergence rates much faster than SGD. Testing this, we found this to hold true for our data training (see Figure 20). Nevertheless, this should be caveated with recent papers which indicate that CNN's trained on SGD might actually perform better on real-world recognition tasks, despite the longer training times and issues with convergence (Wilson, et al. 2017) (Keskar and Socher 2017).

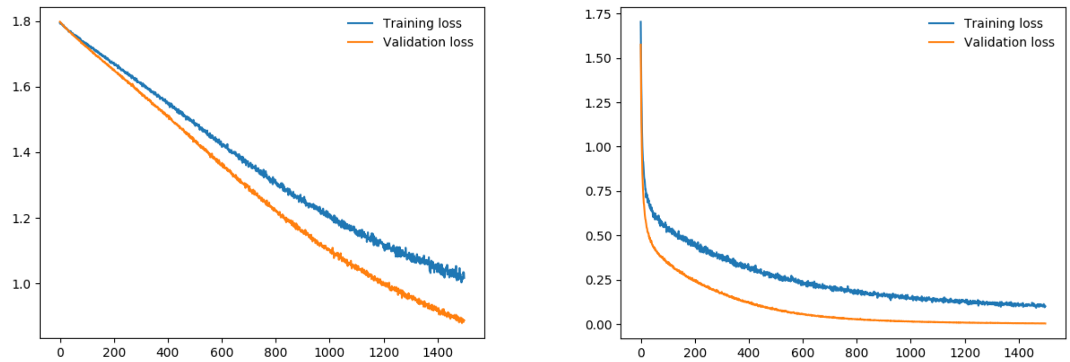

Figure 20 – ResNet101 SGD vs Adam Training and Validation Loss after 1500 Epochs

4.3.4.3 Hyperparameter Selection

In addition to optimization strategy used, there are another set of control factors which are necessary to set when optimizing the neural network. These factors (called **hyperparameters** in order to differentiate them from the weight and bias parameters) determine how the optimizer ingest data, as well as how it will adapt over time. The number of **epochs** for the model to run, the **batch size** of images to be processed in a group, and the **learning rate** (or gradient step size) for adjusting the weights of the network are all elements working in conjunction with the Optimizer algorithm. However, there is no strict mathematical formula for setting these criteria; currently, the advice for determining their value within the CNN community is through previous experience, general rules of thumb, and at best computationally expensive search methods (Rawat & Wang, 2017). For our implementation, we found that a Batch Size of 32 worked on all models (larger numbers resulting in memory errors from the GPU) combined with a Learning Rate of 0.00001.

4.3.5 Computer Setup

Training and testing was performed on a Linux desktop PC with the following technical specifications:

- Ubuntu Linux 64-bit Operating System version 16.04
- Memory 15.6 GiB
- Processor Intel Core i7-4790 CPU @ 3.60GHz x 8
- GPU GeForce GTX 1080/PCIe/SSE2 11GB GDDR5
- Hard Drive Disk 2.0 TB

4.4 RESULTS

	Precision	Recall	F1
VGGNet16			
Test Set 1	0.91798442096	0.916	0.915419496816
Test Set 2	0.753055140463	0.663333333333	0.648033620148
Test Set 3	0.748087773296	0.727741935484	0.703518991967
VGGNet19			
Test Set 1	0.934092753035	0.932	0.931810453649
Test Set 2	0.771980409838	0.703333333333	0.684985977869
Test Set 3	0.720434169913	0.69935483871	0.672505887722
ResNet18			
Test Set 1	0.925838065022	0.924	0.923948956666
Test Set 2	0.78883734633	0.716666666667	0.71025197972
Test Set 3	0.742982665991	0.698064516129	0.691839920569
ResNet32			
Test Set 1	0.949130881915	0.948	0.947767351785
Test Set 2	0.809900926676	0.756666666667	0.751228824163
Test Set 3	0.779995175508	0.765161290323	0.762563231266
ResNet50			
Test Set 1	0.9481757157 21	0.946	0.945695967626
Test Set 2	0.856358294766	0.816666666667	0.813863249005
Test Set 3	0.798356584485	0.789677419355	0.782728035751
ResNet101			
Test Set 1	0.966172508715	0.966	0.965881328159
Test Set 2	0.823854028094	0.776666666667	0.778277926386
Test Set 3	0.780196006414	0.767741935484	0.759586538397
ResNet152			
Test Set 1	0.950859881988	0.95	0.94976110652
Test Set 2	0.824691537068	0.763333333333	0.762530261101
Test Set 3	0.788853057301	0.761290322581	0.75663329502
DenseNet161			
Test Set 1	0.972197265932	0.972	0.971939992201
Test Set 2	0.803469614575	0.773333333333	0.770320699444
Test Set 3	0.804380008612	0.798709677419	0.794285949963

Table 6 - Classification Model Precision, Recall, and F1 Metrics

4.4.1 VGGNet

4.4.1.1 VGGNet16

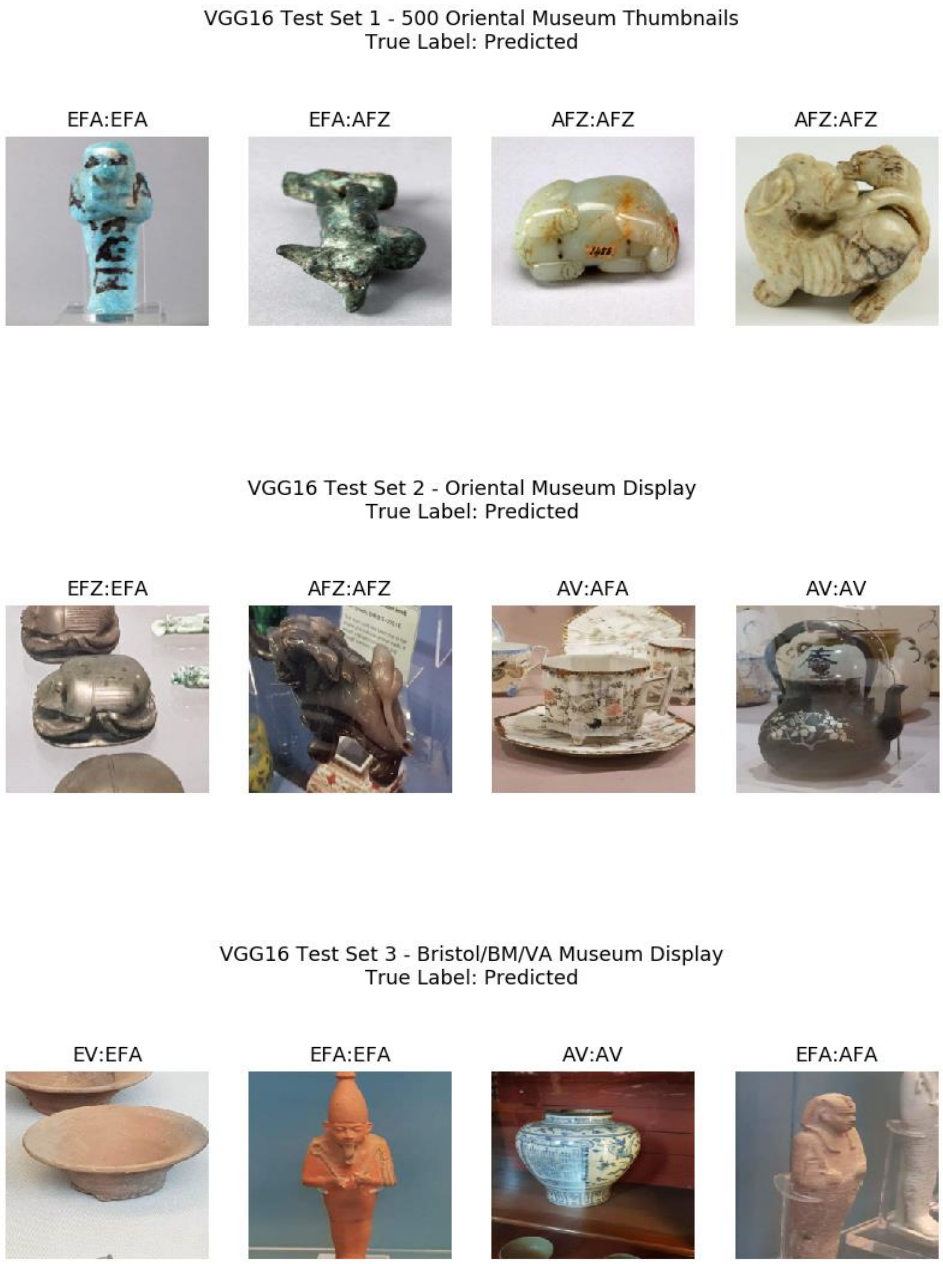

Figure 21 - VGGNet16 Sample Output Predictions (Test Set 1 images owned by Durham University Oriental Museum, Test Set 2 and 3 images taken by Author)

4.4.1.2 *VGGNet16 Confusion Matrix Results*

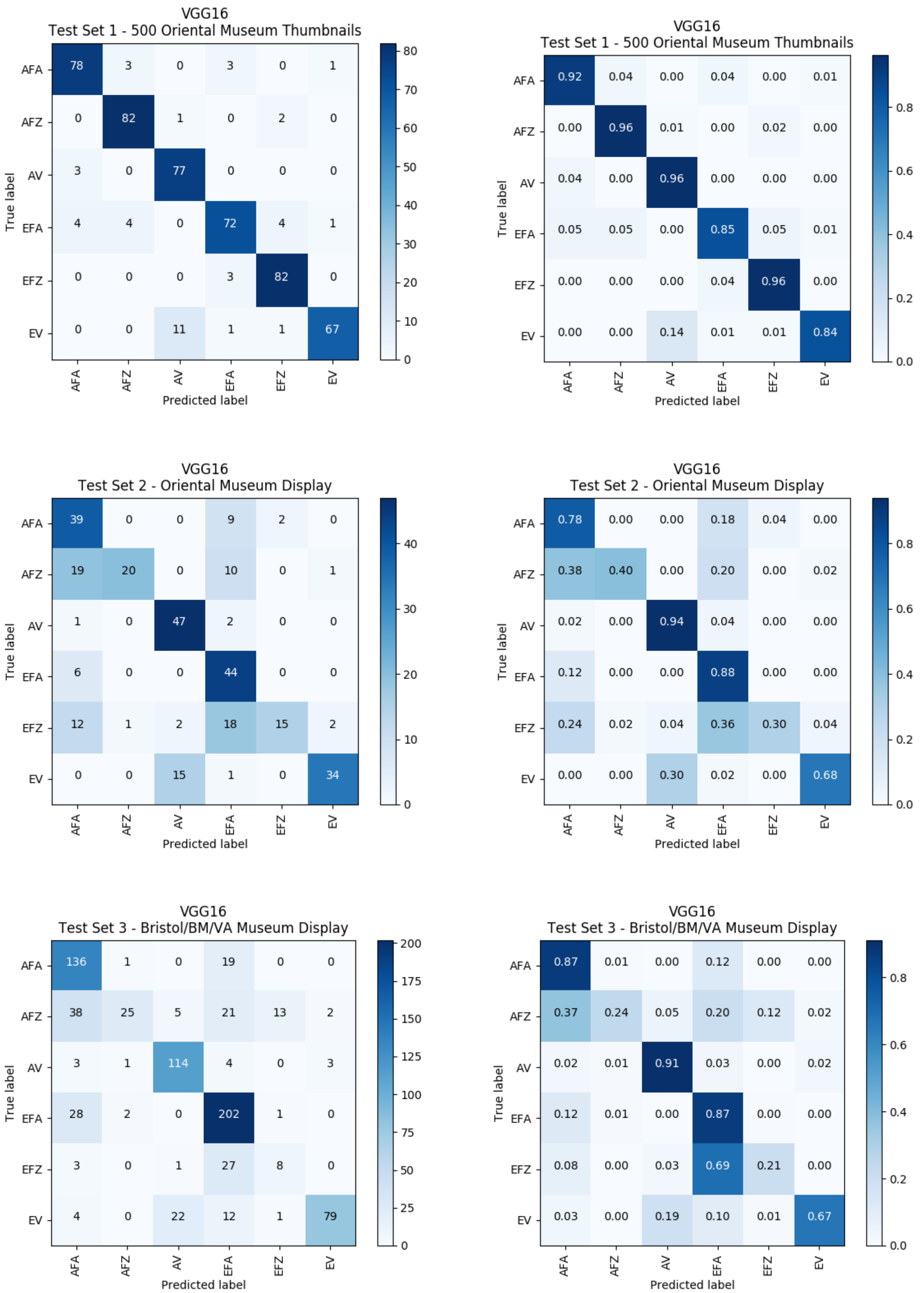

Figure 22 - VGGNet16 Test Set 1, 2, 3 Confusion Matrix Results (Non-normalized and Normalized)

4.4.1.3 VGGNet19

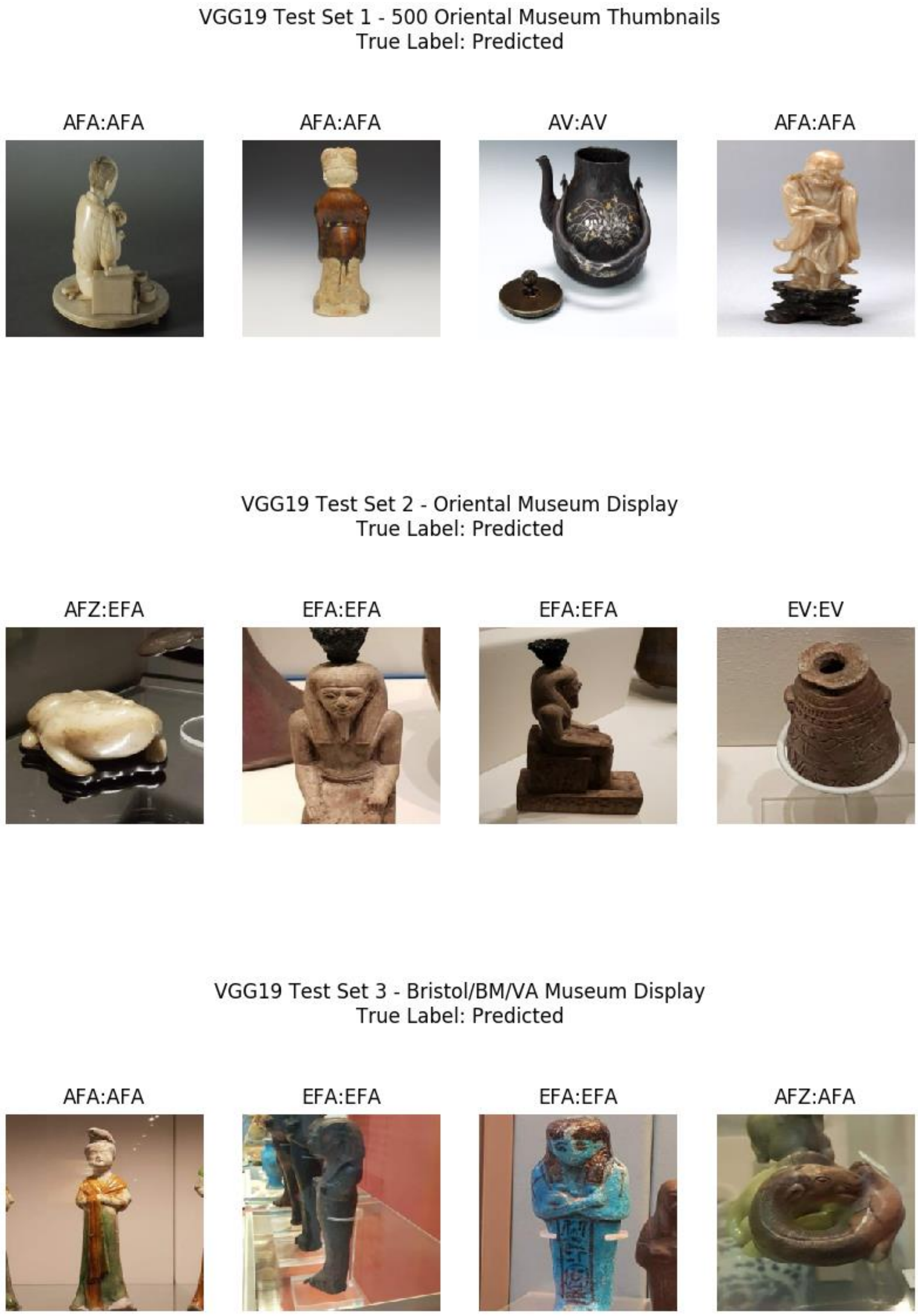

Figure 23 - VGGNet19 Sample Output Predictions (Test Set 1 images owned by Durham University Oriental Museum, Test Set 2 and 3 images taken by Author)

4.4.1.4 VGGNet19 Confusion Matrix Results

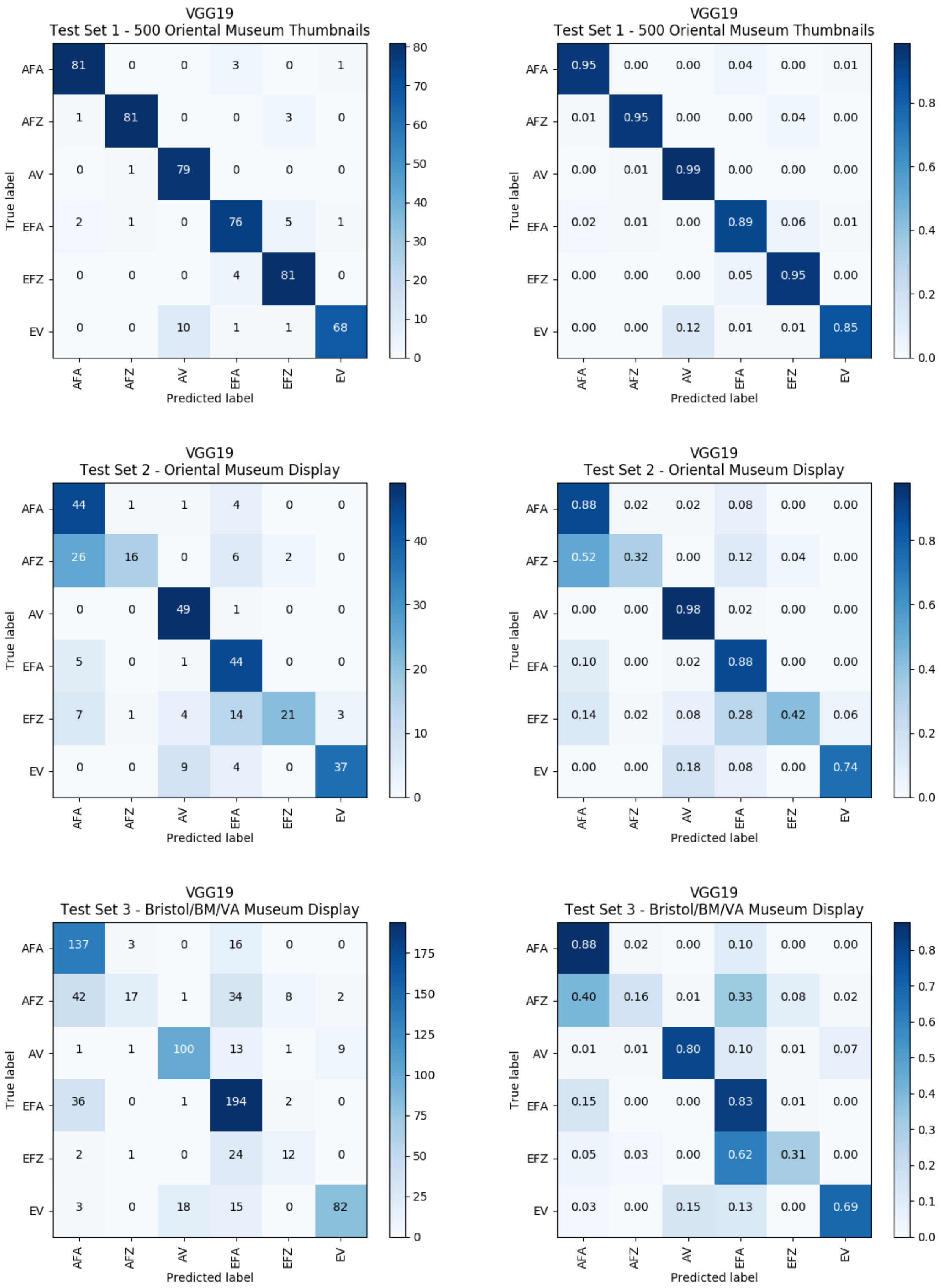

Figure 24 - VGGNet19 Test Set 1, 2, 3 Confusion Matrix Results (Non-normalized and Normalized)

4.4.2 ResNet

4.4.2.1 ResNet18

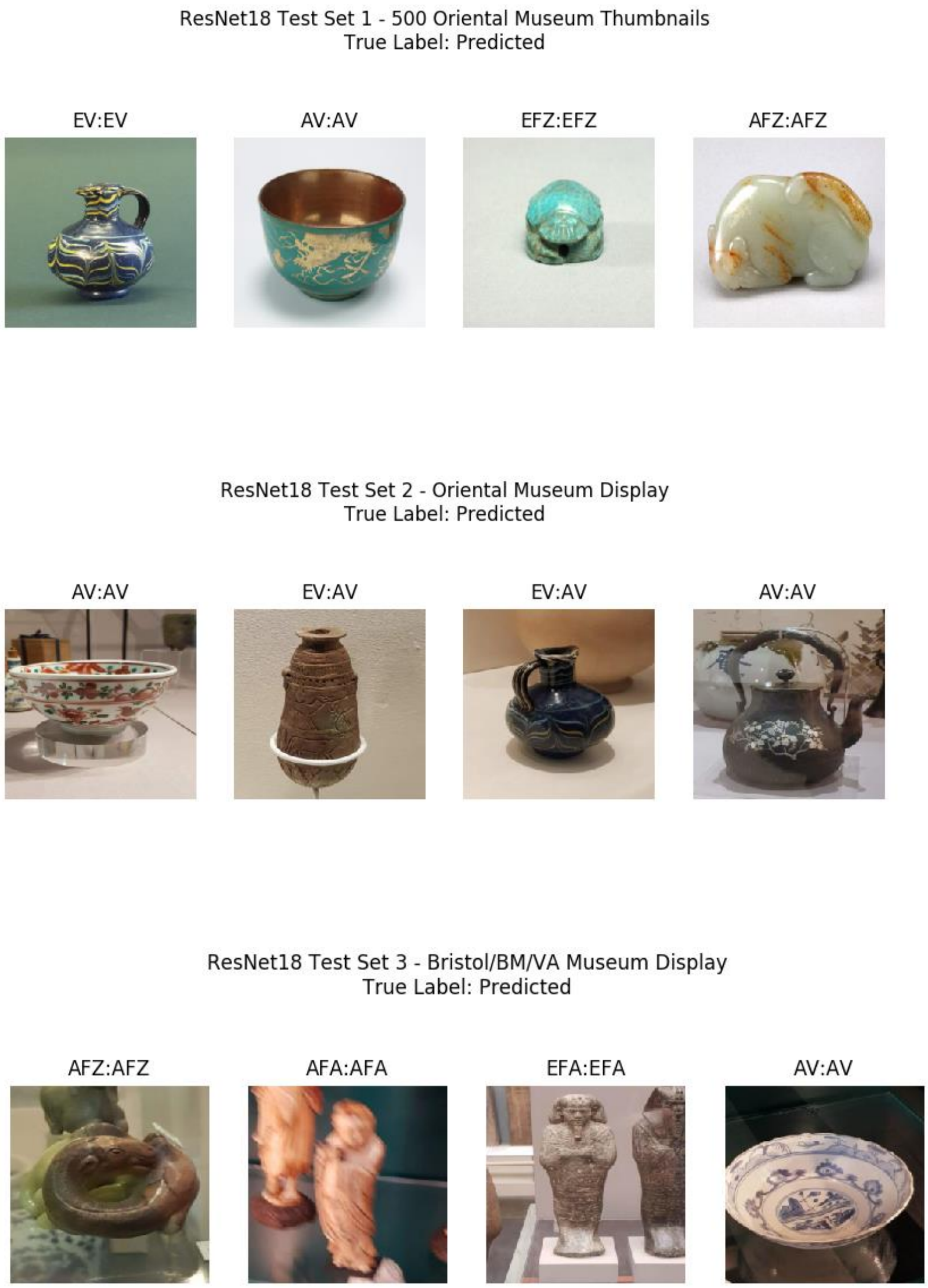

Figure 25 - ResNet18 Sample Output Predictions (Test Set 1 images owned by Durham University Oriental Museum, Test Set 2 and 3 images taken by Author)

4.4.2.2 ResNet18 Confusion Matrix Results

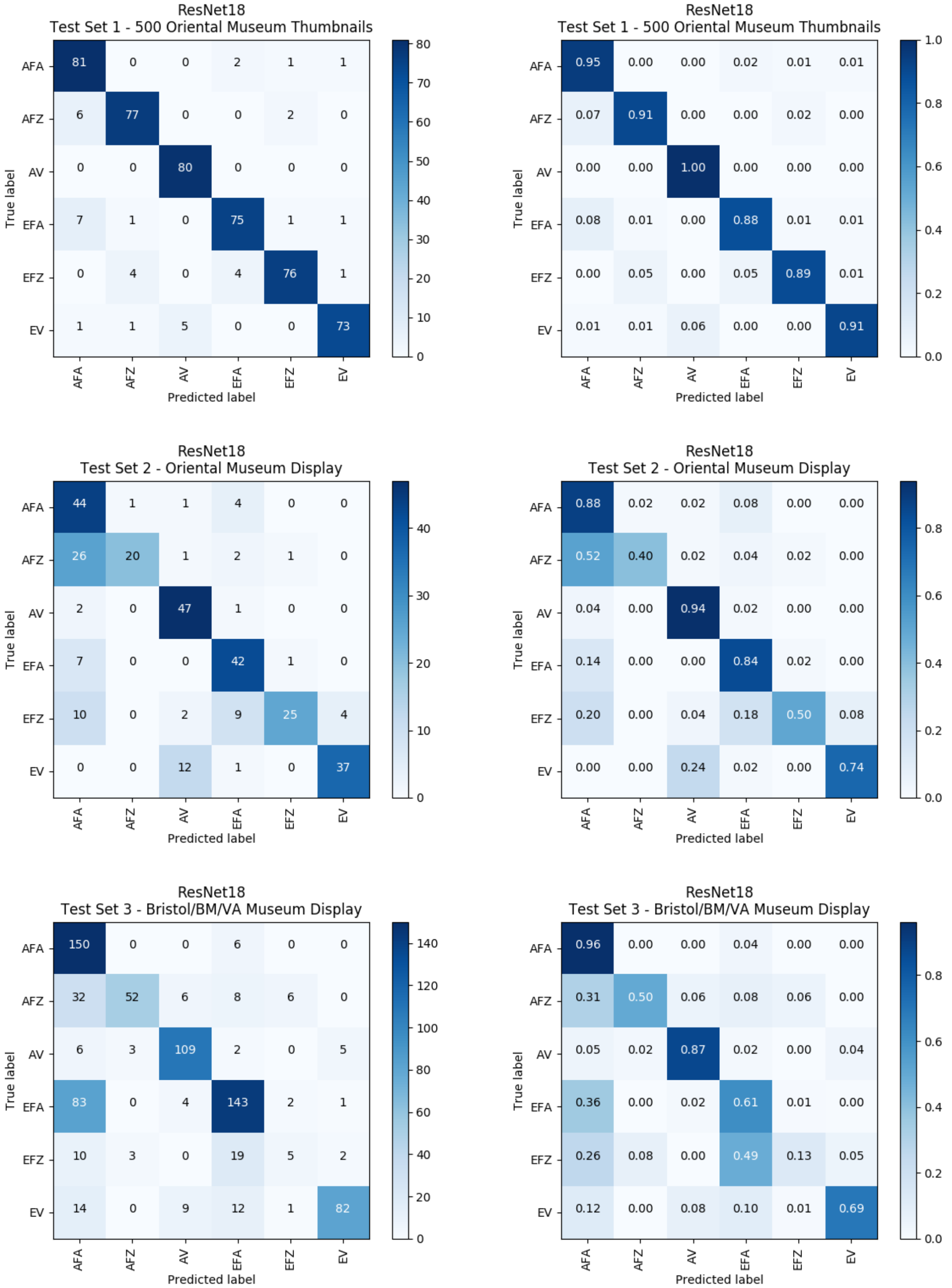

Figure 26 - ResNet18 Test Set 1, 2, 3 Confusion Matrix Results (Non-normalized and Normalized)

4.4.2.3 ResNet32

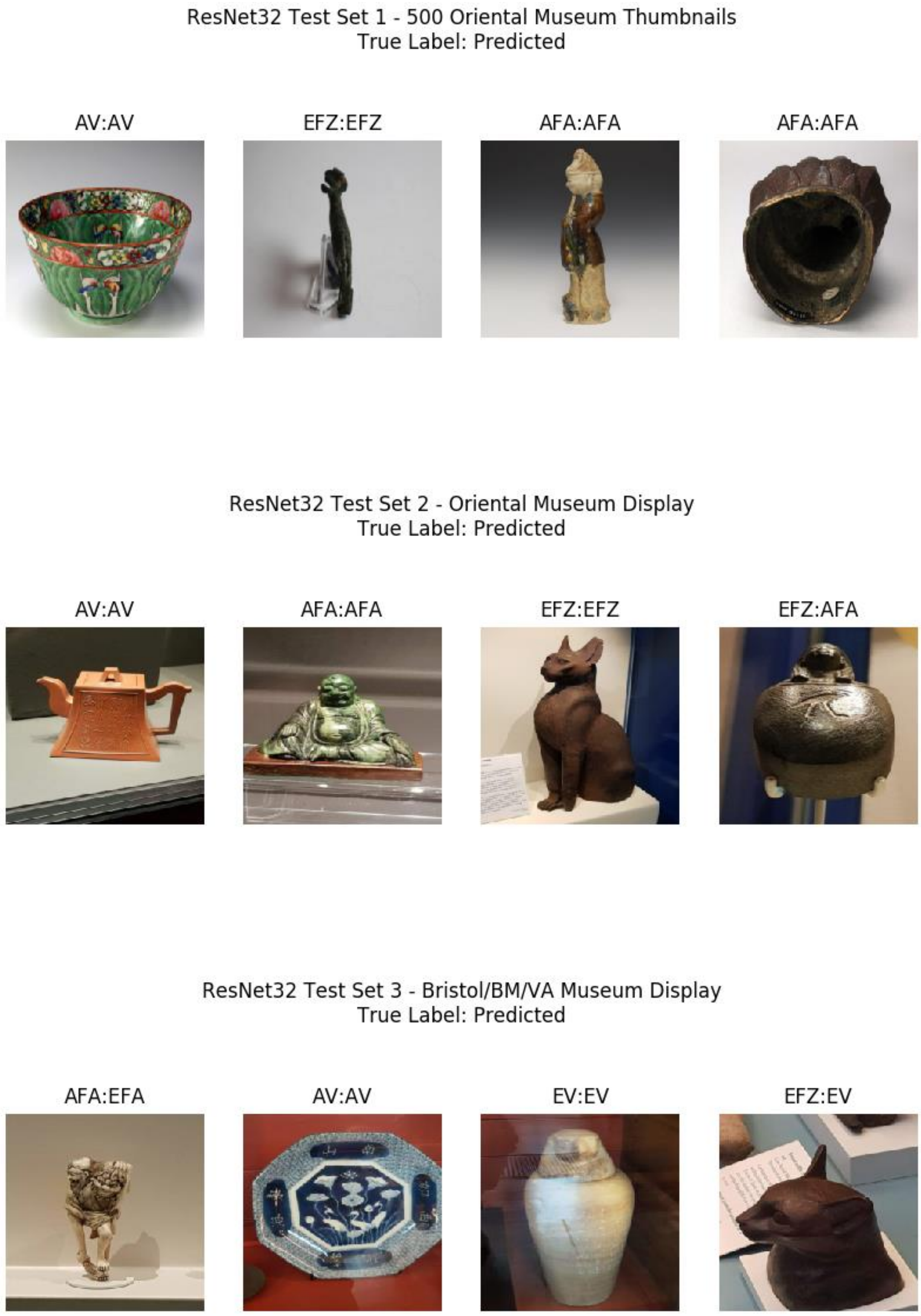

Figure 27 - ResNet32 Sample Output Predictions (Test Set 1 images owned by Durham University Oriental Museum, Test Set 2 and 3 images taken by Author)

4.4.2.4 ResNet32 Confusion Matrix Results

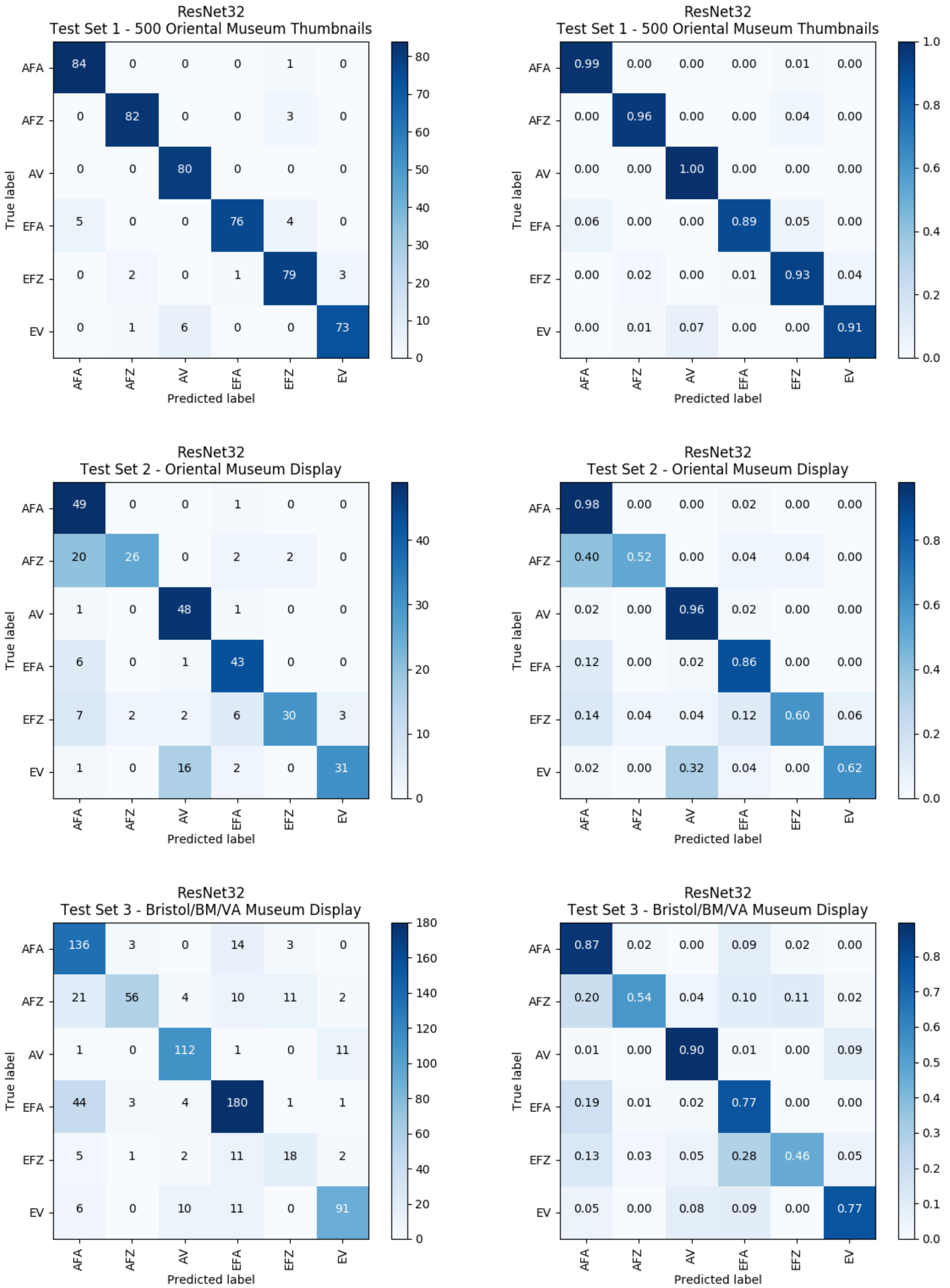

Figure 28 – ResNet32 Test Set 1, 2, 3 Confusion Matrix Results (Non-normalized and Normalized)

4.4.2.5 ResNet50

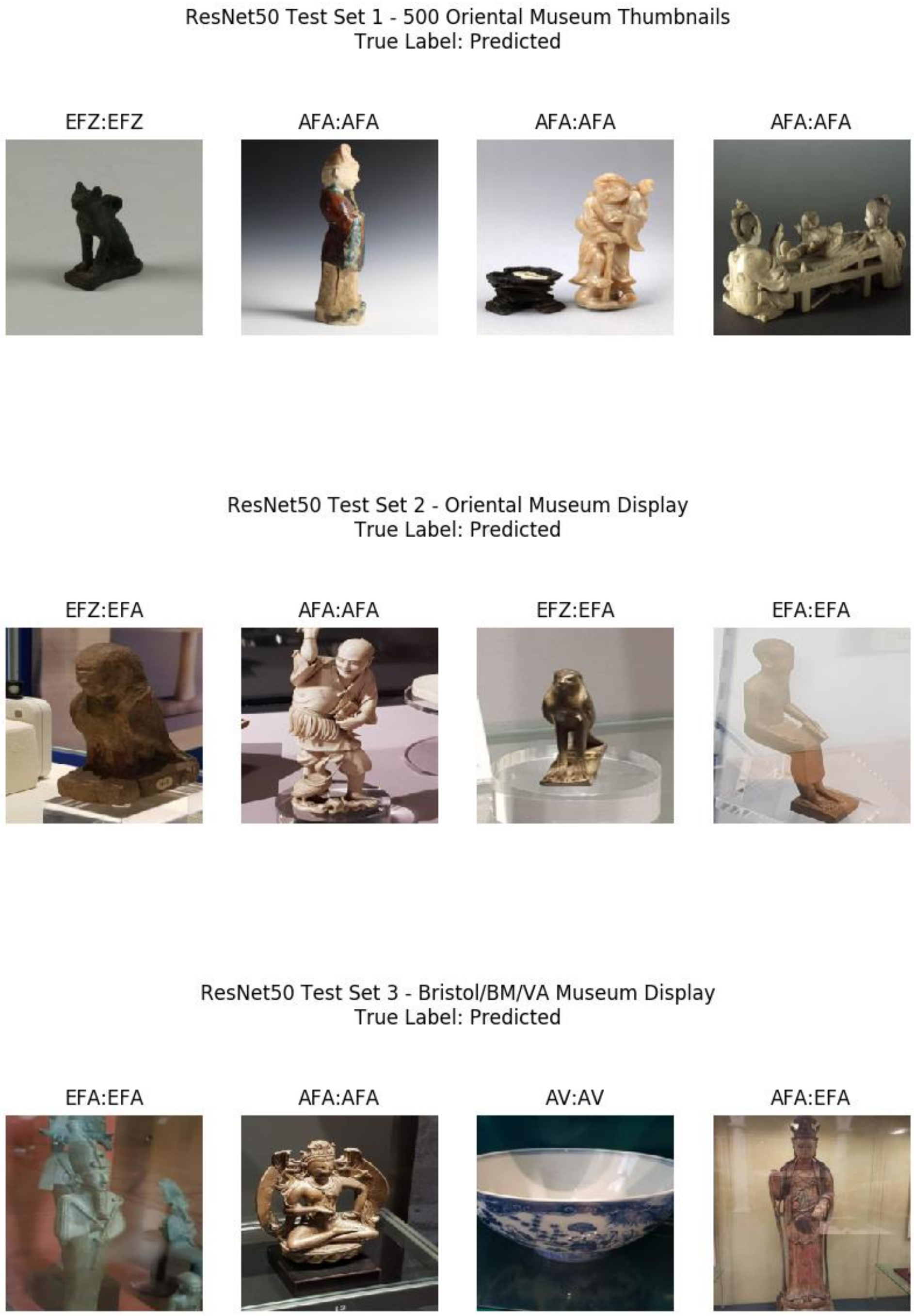

Figure 29 - ResNet50 Sample Output Predictions (Test Set 1 images owned by Durham University Oriental Museum, Test Set 2 and 3 images taken by Author)

4.4.2.6 ResNet50 Confusion Matrix Results

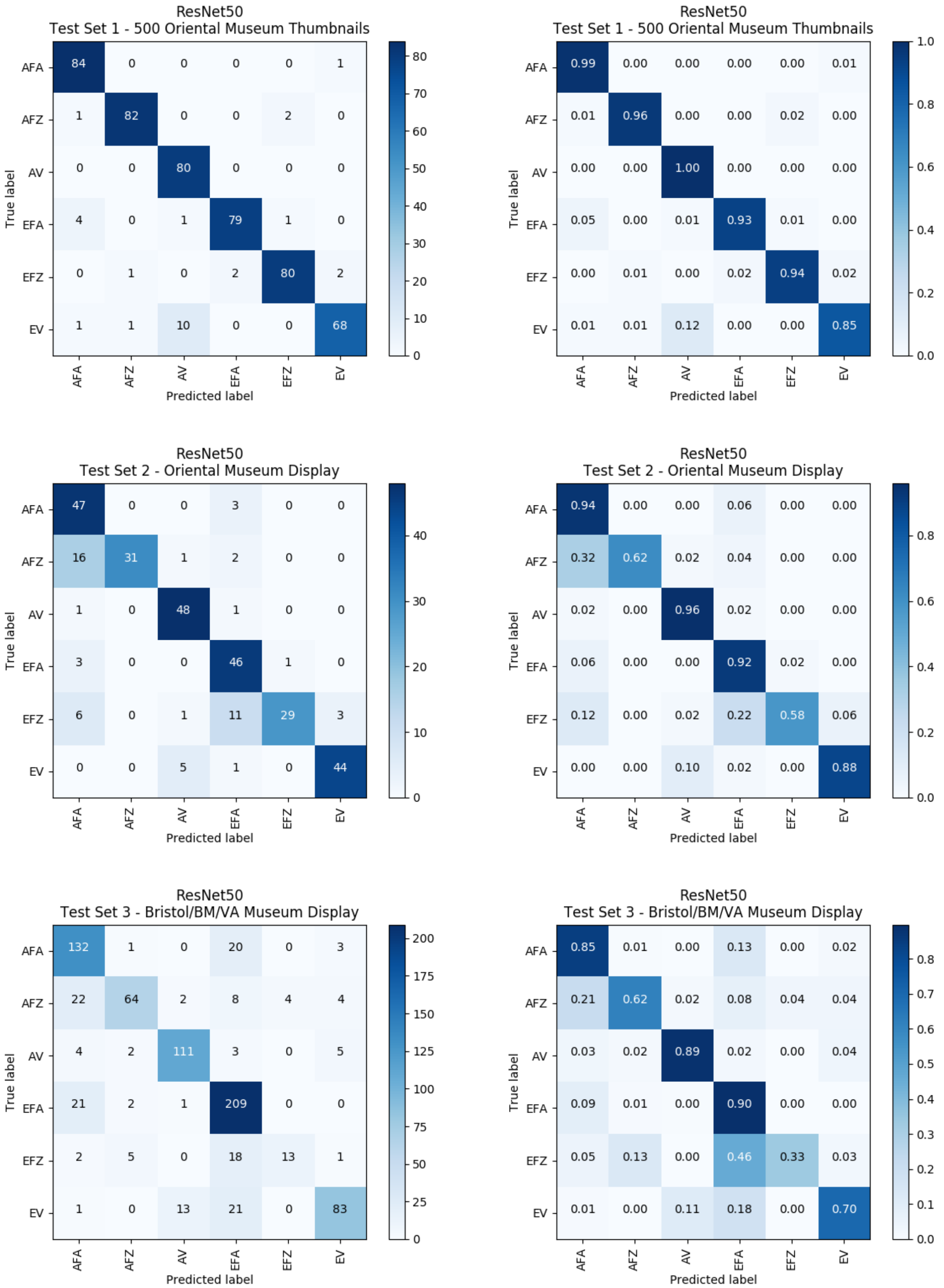

Figure 30 – ResNet50 Test Set 1, 2, 3 Confusion Matrix Results (Non-normalized and Normalized)

4.4.2.7 ResNet101

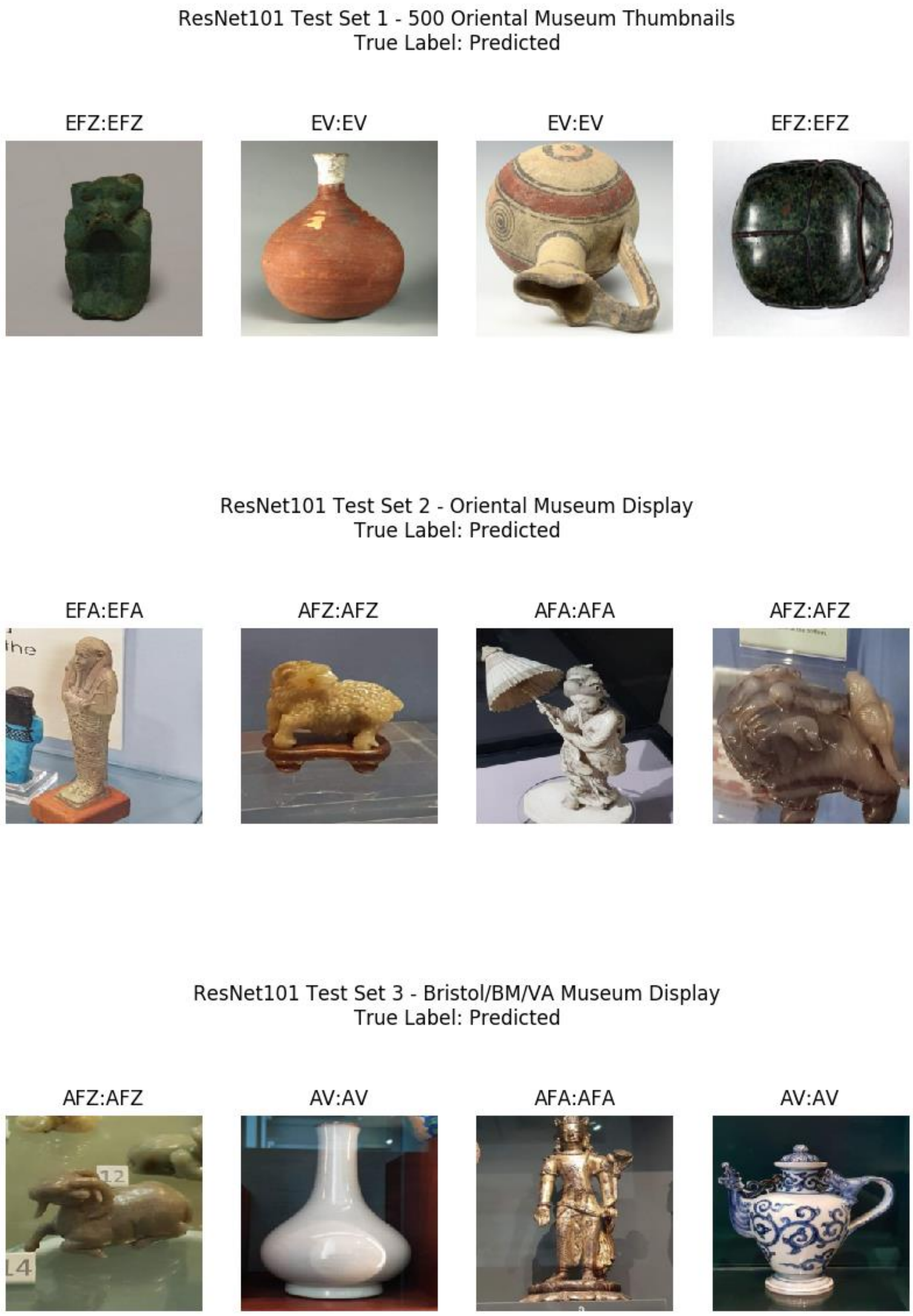

Figure 31- ResNet101 Sample Output Predictions (Test Set 1 images owned by Durham University Oriental Museum, Test Set 2 and 3 images taken by Author)

4.4.2.8 ResNet101 Confusion Matrix Results

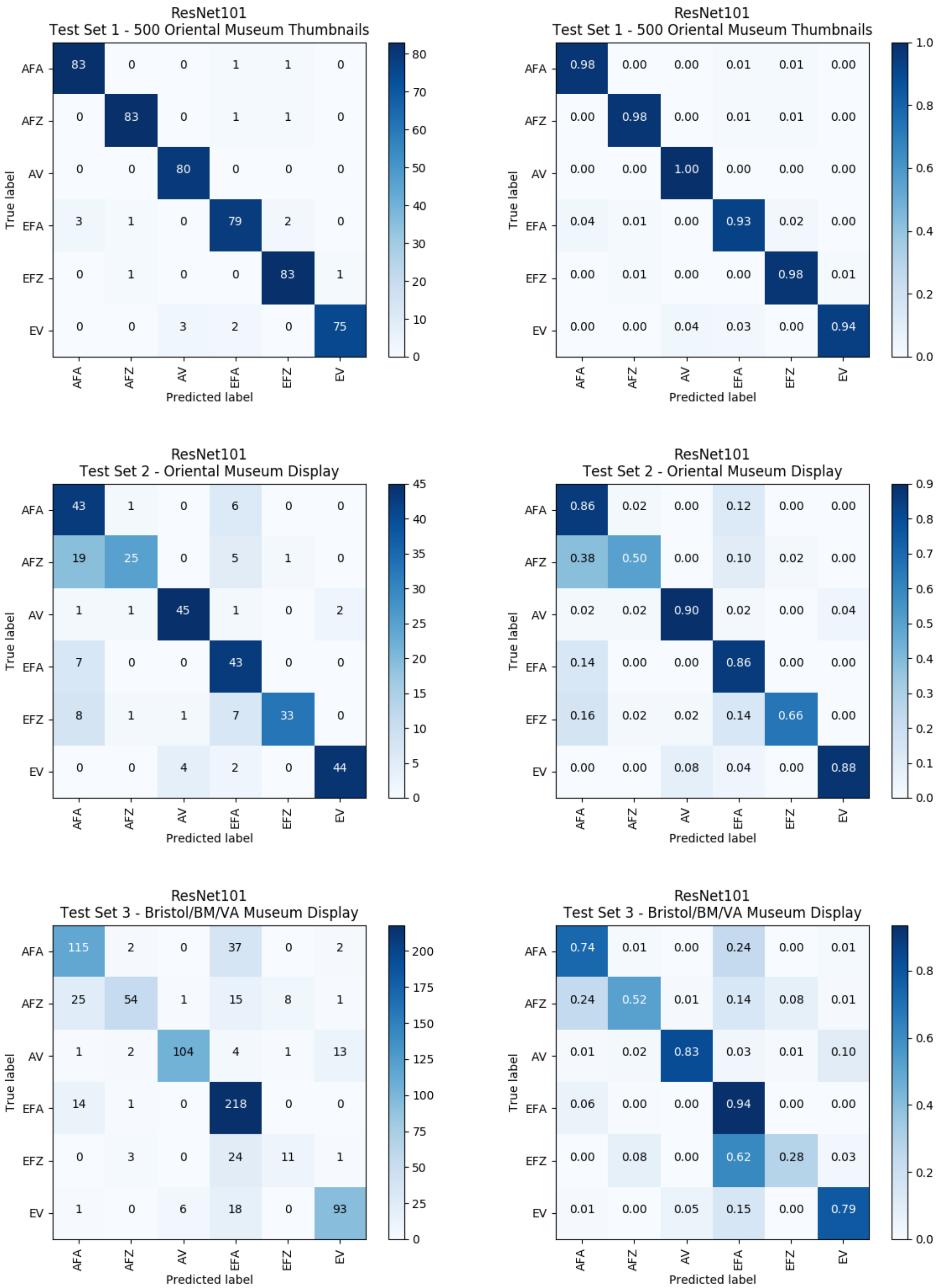

Figure 32 - ResNet101 Test Set 1, 2, 3 Confusion Matrix Results (Non-normalized and Normalized)

4.4.2.9 ResNet152

Figure 33 - ResNet152 Sample Output Predictions (Test Set 1 images owned by Durham University Oriental Museum, Test Set 2 and 3 images taken by Author)

4.4.2.10 ResNet152 Confusion Matrix Results

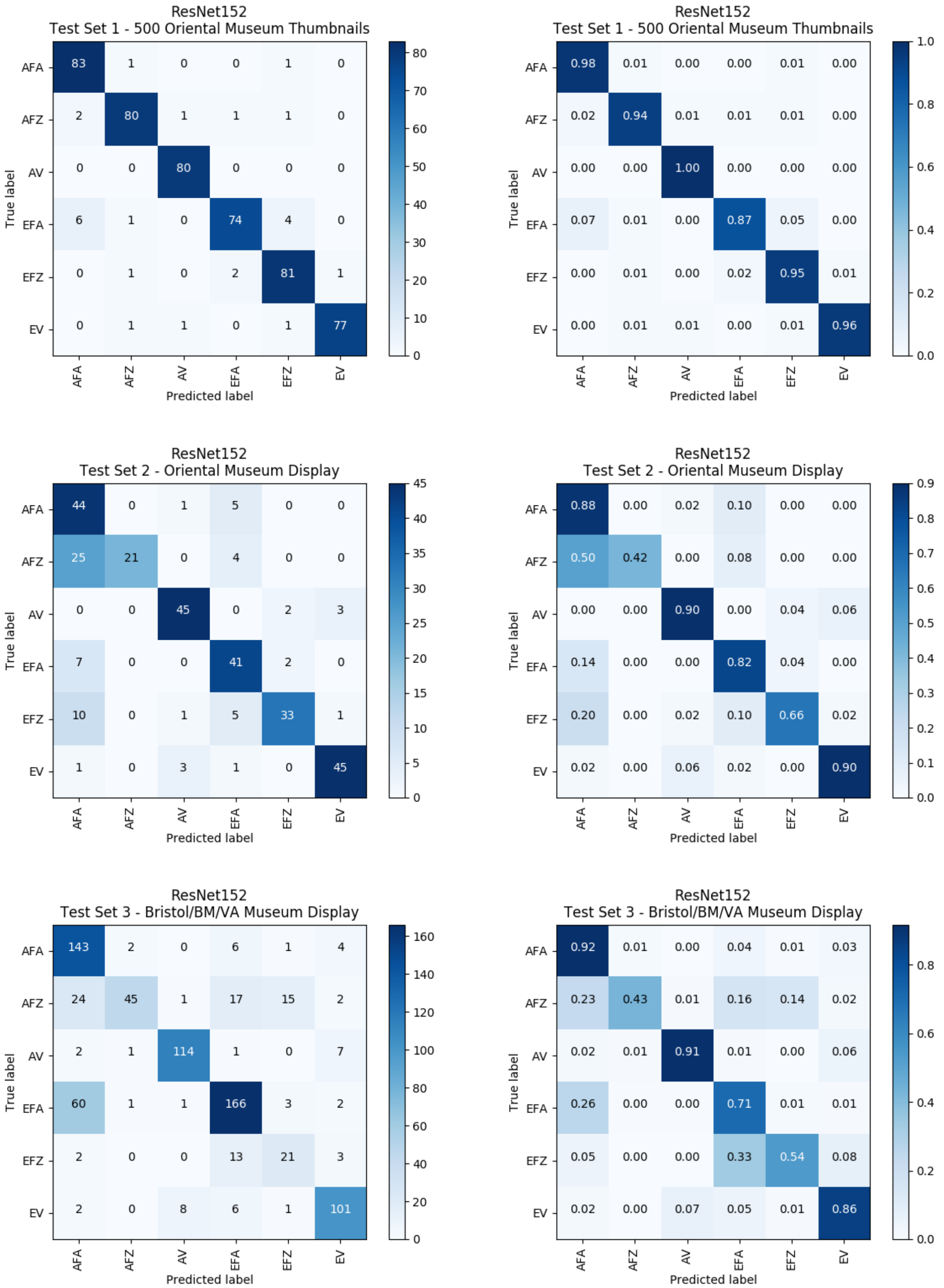

Figure 34 - ResNet152 Test Set 1, 2, 3 Confusion Matrix Results (Non-normalized and Normalized)

4.4.3 DenseNet

4.4.3.1 DenseNet161

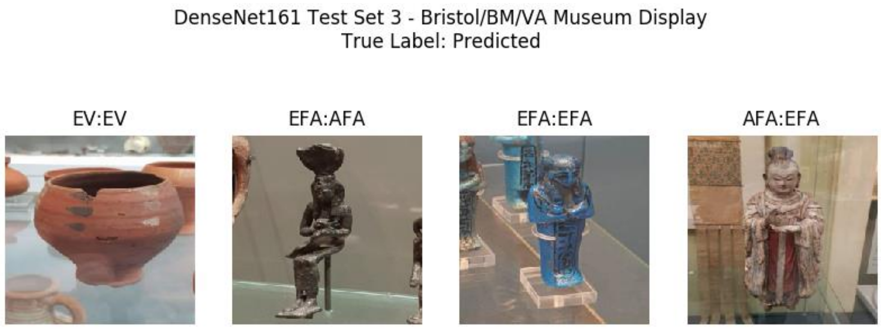

Figure 35 - ResNet161 Sample Output Predictions (Test Set 1 images owned by Durham University Oriental Museum, Test Set 2 and 3 images taken by Author)

4.4.3.2 DenseNet161 Confusion Matrix Results

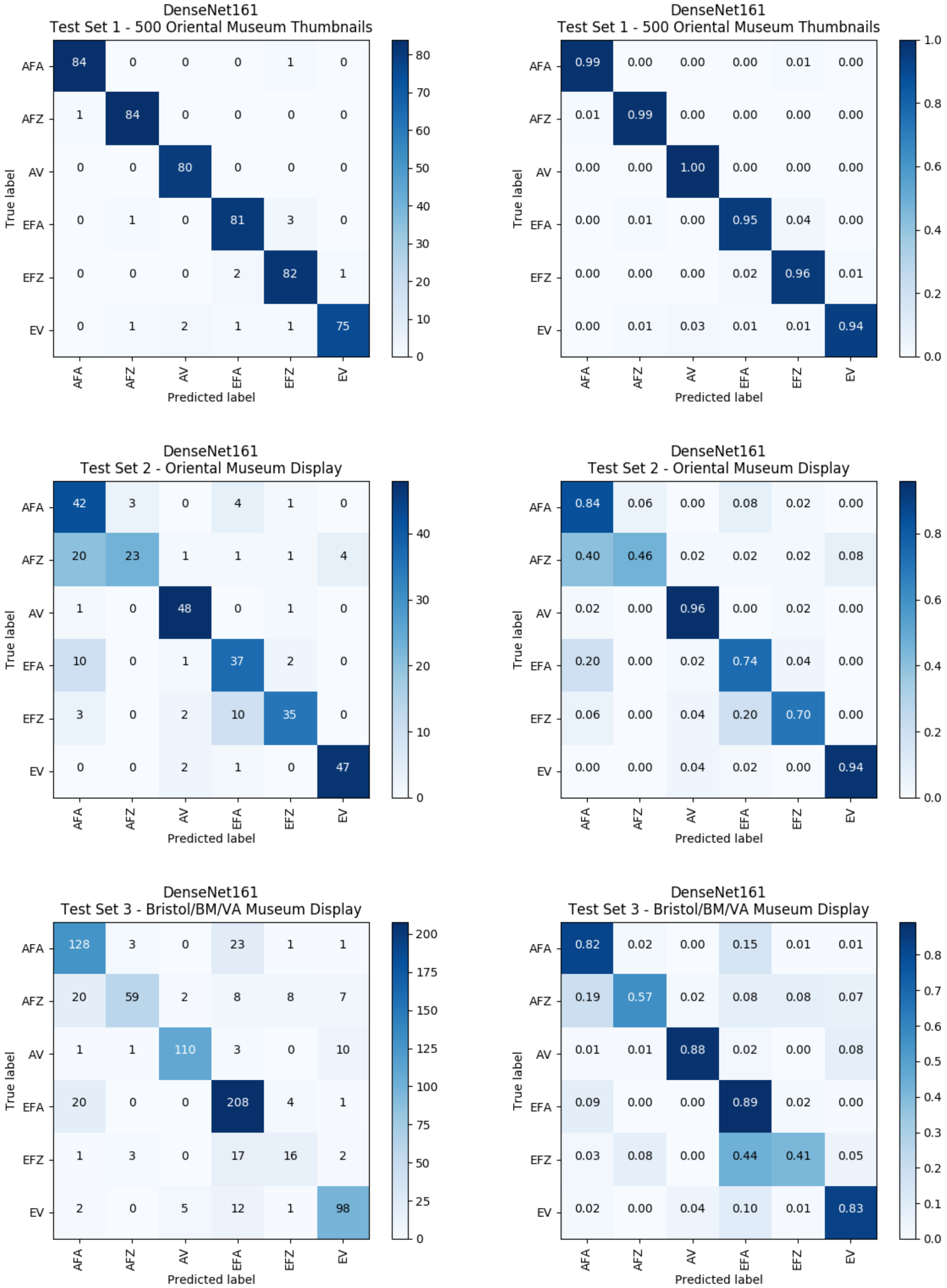

Figure 36 – DenseNet161 Test Set 1, 2, 3 Confusion Matrix Results (Non-normalized and Normalized)

4.5.1 Epoch Training Times

The VGGNet architectures achieved 100% Validation Accuracy within the first 100 epochs. Training for the ResNet architectures was done up to 1500 epochs. ResNet18 and 32 generally took the longest number of epochs to reach 100% Validation Accuracy, in excess of 1000 epochs. The deeper architectures (50, 101, and 152) achieved 100% Validation Accuracy within 1000 epochs. Similarly, DenseNet achieved 100% Validation Accuracy within the first 1000 epochs.

4.6 Discussion

For the overall results, the model performances showed prediction accuracies tied the model structure:

- **Test Set 1**: In the base case of Classification, all of the models achieved over 90% classification accuracy according to the performance metrics. DenseNet161 scored the highest in Recall and Precision, achieving a combined F-Score of 97% prediction. For the most part, for Test Case 1 the deeper the layers the better the performance; ResNet152 is the exception, only achieving 95% accuracy as compared to ResNet101's 96% (similar to ResNet50).
- **Test Set 2**: In the first of our challenging cases, classification accuracy dropped off. The best metric performance came from ResNet50, achieving a Precision score of 85% and Recall of 81%. The shallowest models (VGGNet16, 19, and ResNet18) reported the lowest performances.
- **Test Set 3**: Reported metrics were similar to those for Test Set 2, though slightly less accurate. In this instance, DenseNet161 reported the best performance, with an F score of 79% over ResNet50's 78%. As with Test Set 2, the shallowest models performed less well.

Overall, the results demonstrate that the best performance for the training data comes from a controlled setting, with the same objects photographed against the same background, lighting, and angle. While potentially useful for certain heritage applications, for our primary research problem the more significant issue is how well the trained models perform when challenged with images developed based on common recognition challenges that could be encountered with a trafficked object. In this case, the deeply layered models still perform well on objects contained within the training data (76 to 85% Precision and Recall scores), though classification of non-trained objects shows a decline is prediction accuracy (73 to 79% scores).

Studying the Confusion Matrix graphs, it appears that the models have the most difficulty with the Zoomorphic classes in the display case tests, both for Egyptian and Asiatic groups. The Asiatic Zoomorphic Figurines in particular became less certain in Test Set 2 and 3. Re-examining our class composition (Figure 9, Figure 12), the wider variety of feature morphologies seems like one possible reason for this performance issue.

4.7 Conclusion

Continuing from the previous chapter, we can now start to draw some conclusions on how well the application of machine learning to artefact typology classification works when developed from a museum's digital records. Given our data set size we chose to utilize transfer learning as our neural network training framework, a type of machine learning commonly used in the data science community when the image sample size is limited. We developed a specific implementation using the open source PyTorch development tools and libraries, based on experimentation with image processing and data loading, optimization functions, CNN models, and hyperparameter selections. The final result was experimental results which consistently achieved 100% validation accuracy in training.

Next we evaluated the trained models using three variations of test sets. Using the evaluation metrics and confusion matrix visualization described in the previous chapter, we see that CNN Models of at least 50 layer complexity provides the highest accuracies. The test set composed of digital data created using a controlled environment (the Oriental Museum's photo lab) results in the greatest accuracies, regularly achieving greater than 90% predicted type classification. The challenging test sets saw a steeper drop-off; however we still see a strong central line in most of the confusion matrices, with the zoological figurines (specifically Asiatic) producing the weakest results.

Considering our main research question, overall the results show promise, though not without several caveats. We can conclude that a museum's digital records can work well as training data for heritage objects if they have been consistently photographed and labelled. Further, broad classes and categories created based on similarity and repeated features will achieve high prediction rates. However, care should be taken to not overextend class types, as an overabundance of different varied forms (even those which share the same artistic style and material) can degrade performance. In this iteration one of the main difficulties was in class creation, with much of the effort geared towards trying to strike a balance between typology groupings with the same number of total images without creating too broad a category. Based on the results, it appears that while the training classes of 700 - 900 images each did produce consistent predictions they begin to fail at broader applications. For the next iterative phase of development, reducing and homogenizing the training class numbers should be a key element of experimentation. This can be paired with the next research question, exploring the use of our developed model in service to Object Identification.

5 AUTOMATED RECOGNITION: OBJECT IDENTIFICATION

Within the field of Computer Vision the process of detecting and labelling a subject within a digital image is grouped under the generic term of Object Recognition. In developing/testing our data set and model, we chose broad classes of object descriptions as our label training sets (combining artefact typology and provenance) similar to the categories of information used by international standards to document and describe an artefact (e.g. ObjectID). Doing so, we quantified the performance accuracy on real-world test sets using our recognition metrics. Returning to the original problem definition, our next step is to consider another application for this system, to see if the developed software can fulfil a defined system solution related to the trafficking of illicit antiquities.

5.1 HERITAGE DOCUMENTATION AND PROTECTION SYSTEM

The Training in Action project[6] is a collaboration between several stakeholders concerned with current issues pertaining to the illicit traffic of antiquities in the Middle East and North Africa (MENA) region. They have a specific focus on Libya due to the ongoing conflict there (Guerin 2019), along with Tunisia as a coordination and training location. The program seeks to train staff in these regions with the documentation techniques necessary to preserve their cultural heritage, as well as providing them with tools for custodianship and protection. The Heritage Documentation and Protection (HeDAP) system is one such development from that effort, with four central priorities:

1. Establish a shared protocol for the documentation and recording of archaeology and portable objects in MENA countries
2. Implement modern methods, techniques, and tools for the rapid recording of at-risk cultural heritage material in MENA countries
3. Develop sustainable expertise within local responsible stakeholders
4. Coordinate with the international community by developing a sustainable system of communication and information sharing

[6] https://www.britishcouncil.org/arts/culture-development/cultural-protection-fund/projects/training-in-action

To achieve the above, a combination of software and hardware tools have been developed for deployment in Tunisia and Libya. A core component is the HeDAP Android tablet app, intended as one part of the solution achieving the above goals. It has been developed in two phases: in Phase 1 the initial prototype was created at Durham University (with support by the British Council), and was field tested in both Libya and Tunisia. It is currently undergoing Phase 2 of development by the L-P Archaeology commercial archaeology partnership, who specialize in digital development of tools intended for fieldwork services[7]. This effort includes developing three primary components for the HeDAP system: the Android App for artefact recording, a database intended to provide long-term storage of data collected from the app, and a portal intended to host HeDAP database instances and provide user support channels. For our research the Phase 1 HeDAP Android App (running on a Galaxy tablet) was available, while the Phase 2 software system was still in development.

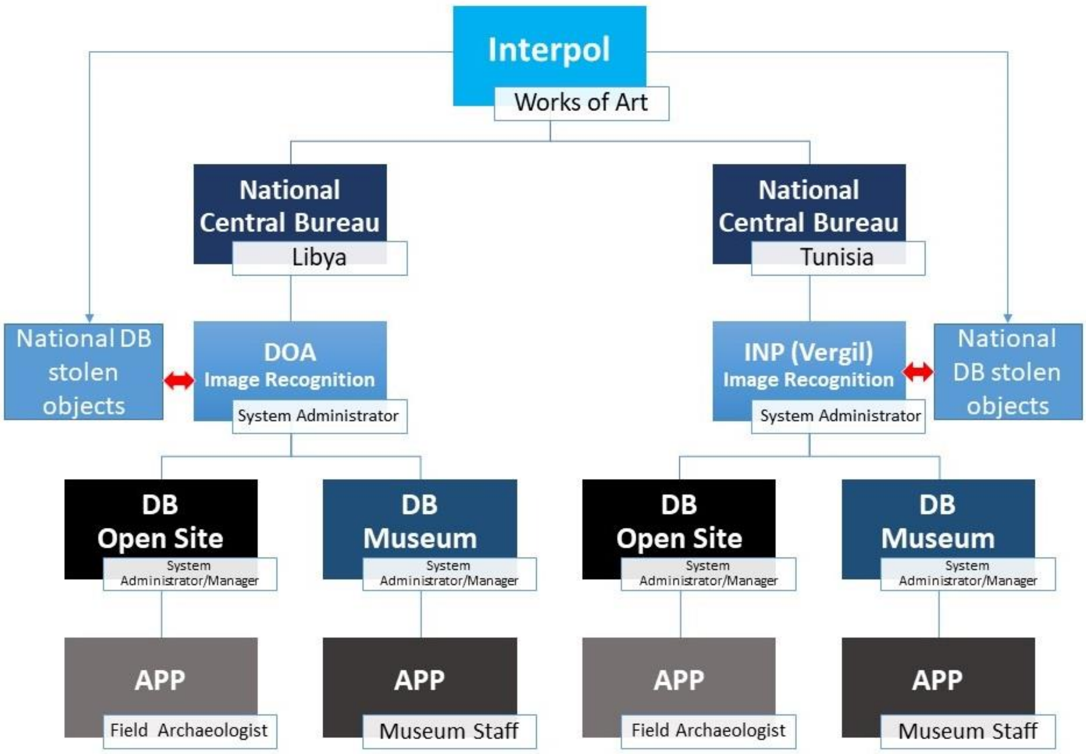

Figure 37 - HeDAP Notional System Diagram, source Professor Anna Leone

The conceptual system (see Figure 37) further seeks to extend the capability of these components, combining them with a robust computer vision system intended to process the

[7] https://www.lparchaeology.com/about/history-of-lp/

image data collected from the app. In the proposed system design, an image recognition system would be layered atop the image database and portal produced in Phase 2. In this system design, the digital image records are intended for use as proof-of-provenance, and as insurance against possible future looting. The desire is for a computer vision system which can analyze a digital image of an object taken by a heritage professional (who suspects it of possibly being looted), and automatically match it to an entry within the HeDAP database. In effect, the goal is for a tool similar to a thumbprint or facial recognition system, dedicated to cultural antiquities from MENA regions.

To provide such a solution is not precisely Image Classification, but more specifically Object Identification. Our question, then, is whether the software developed thus far by our research can be re-purposed to fulfil this need. To do so, we first need to consider our training data.

5.2 Material Background

5.2.1 At-Risk Heritage Collection Evaluation

In addition to supporting the widespread adoptions and training in ObjectID, ICOM is a global leader in coordinating efforts against illegal antiquities smuggling (Howarth 2012). Their Red Lists (as detailed in Section 2.3.3) describe the types of antiquities which are considered most endangered from social/political instability and illicit market trends, by identifying a number of possible categories of artefact considered at-risk. Initially our hope had been to use digital images of material from Libya provided through field usage of the HeDAP app, potentially focusing on those objects highlighted in the Red List of Libyan Antiquities. Because of Libya's historical context, this includes a number of both Graeco-Roman and Islamic era material: Sculptures and Reliefs, Architectural Elements, Instruments, Accessories, and Coins are all potential targets for illegal acquisition. However, direct access to this type of image data was not initially available during the research period. As a possible secondary approach, we contacted the Durham University's Museum of Archaeology regarding their Classics sculpture material, to potentially use similar sculptural material as stand-in for at-risk antiquities from Libya's Roman period (Mugnai 2017). Unfortunately, discussion with the curator and a collections search found that the number of such items was too few, and in many cases specific to local Roman Britannic traditions.

We opted instead to continue working primarily with the Oriental Museum's Egyptian collection. Though not directly related to the current HeDAP area of operations, Egypt remains a primary target of the modern illegal antiquities market, and the Oriental Museum's collection continued to provide us with the greatest flexibility in terms of access to their material resources.

5.2.2 Egyptian Figurine Single Object Data Set

The choice was made to focus on a relatively homogenous set of figurines from the Egyptian collection to create our single object classes, in order to test how well several objects of the same 'type' (i.e. with similar material, dimensions, and overall shape) could be differentiated from one another. Working with the Oriental Museum's collection system, a set of nine figurines were selected for Identification: three bull figurines, four Osiris figurines, a cat figurine, and a Khnum figurine (see Appendix I Wellcome Egyptian Figurines). By including two categories of similarly shaped objects (along with two unique figures), we hoped to test the capability of the model to differentiate between individual structural elements, even when the overall design is similar.

5.3 Data Preparation and Modelling

5.3.1 Digital Image Capture

Our previous data labels were based in part on the categories contained within heritage standards, broadly describing the typology and origin of the object. It was developed from the method and manner of documentation employed by the Oriental Museum for their digital imaging, and benefited from the fact that their data had been collected over the course of a decade (with multiple perspective digital photographs of each object).

In developing potential methods for a new training data set, one initial consideration was the use of Photogrammetry and 3D image models as a way in which synthetic images could be generated for a single object. While ultimately that form of modelling was rejected as not being feasibly within the scope of the HeDAP app's proposed design requirements, it did provide a conceptual inspiration for approaching the problem of single object identification in our system. Photogrammetry uses multiple images to construct a three-dimensional mapping of some target, with workflow instructions for how those images should be acquired in order to construct the digital simulacrum. If we follow that same line of thought, a similar type of data collection could be used with our established CNN model training software. Many modern digital cameras have a mode for rapidly taking multiple images in succession, general referred to as Burst Mode Photography. Traditionally, this mode has been used to capture a subject in motion, with the continuous recording a way to avoid blurring (Grigonis 2019). Early on the number of photos taken were limited, often only 10 or 20 shots at high resolution. However, improvements in hand-held mobile devices data storage have increased the number of images taken in a photoburst: a Galaxy S7 Android phone, for example, is capable of taking 100 continuous image shots at a time. Capitalizing on this capability, it is feasible to create a digital documentation workflow similar to the way in which photogrammatic image capture works, to record several hundreds of images of a single object. These images can, in turn, be used as the training set of images for our CNN model, with each single object acting as the "class" that the model is attempting to develop a data mapping for (see Figure 38). The Oriental Museum's photography lab was used, with the same background, lighting, and object stands used for the DSLR image capture described in Chapter 4.

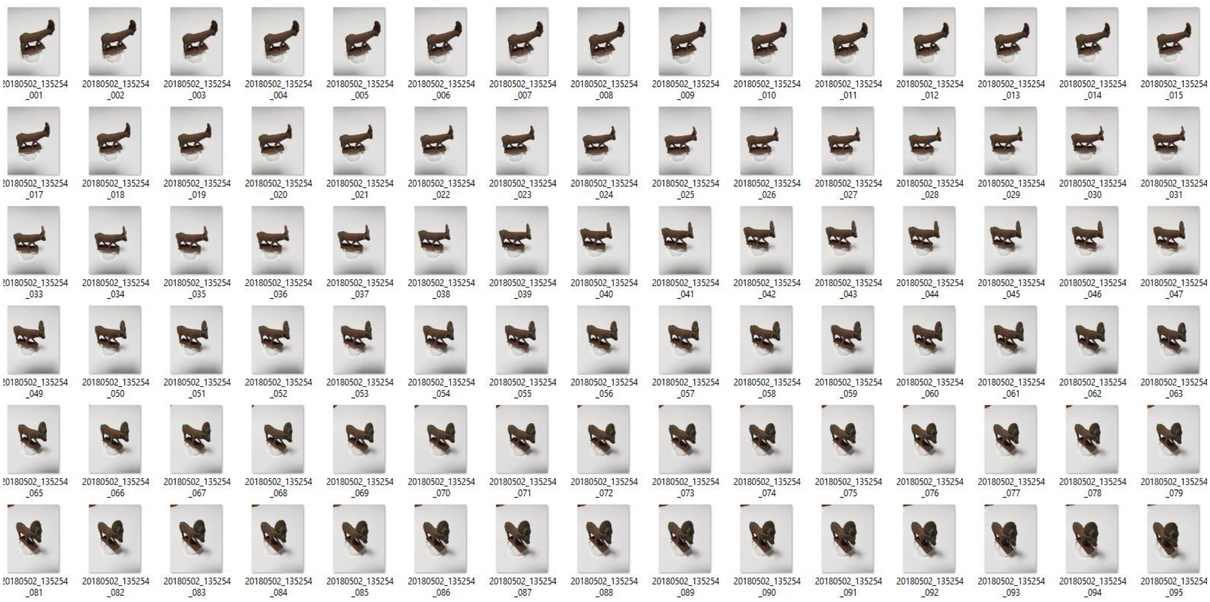

Figure 38 - Sample of photoburst images taken as training input for Object Identifiation, of Hathor Bull figurine (DUROM.1971.105). Source Author.

Burst Mode Image capture was done using the Galaxy S8 camera used for the previous data recording. Additionally, an Android Tablet with the HeDAP app installed was used for Test Image data collection. For Burst Mode, the object would be positioned in the center of the recording space, and a single 100-shot burst taken moving the camera in an arcing motion from left to right. After a single burst shot was completed, the figurine would be turned slightly, and the photography repeated in the same manner. Once all sides had been photographed, a final photoburst arc was done moving over the top of the figurine. Similar to the single shot photographs, this method was done in order to capture the relative front, back, sides, and top of the figurine. The HeDAP photography, meanwhile, was done following a similar pattern as used with the DSLR camera, focusing on capturing front, back, left, right, and top images. For most of the objects 400 photographs were taken; EG983 (and Osiris figurine) and EG1424 (the cat figurine) only had 300 photos taken, due to a recording error. For the HeDAP test set, most of the objects had 5 to 8 images taken, though EG943 includes twice that number due to a second set of shots being taken from a greater distance away (included to check if proximity would affect accuracy). In total, the training set contained 3,400 images and the test set 69 (see Table 7).

Accession Number	**Short Description**	**Burst Mode Images**	**HeDAP Images**
DUROM.1971.105	(bronze) Bull Figurine	400	7
EG943	(bronze) Apis Bull Figurine	400	14
EG983	(bronze) Osiris Figurine	300	7
EG999	(bronze) Apis Bull Figurine	400	6
EG1424	(faience) Cat Figurine	300	5
EG1493	(bronze) Khnum Figurine	400	8
EG1495	(bronze) Osiris Figurine	400	7
EG1614	(bronze) Osiris Figurine	400	8
EG6141	(bronze) Osiris Figurine	400	7
		3400	**69**

Table 7 - Single Object Image Capture

5.3.2 Model Structure and Training

For our CNN models we kept the same data regularization, loading, and augmentations. The same set of image transformations were performed (random vertical and horizontal flips, rotations up to 180 degrees, brightness variations and random conversion to grayscale), with a 80% Training and 20% Validation data loading split. VGGNet19, all variations of ResNet, and DenseNet161 were retained as our selected models with the same backpropagation architecture (Negative Log Likelihood Loss as the loss function and Adam as the optimizer). For the hyperparameter selection, batch size was kept at 32 images per processing group. Checking the Training and Validation Loss graphs (along with the Validation Accuracies) demonstrated that training on the new data set was converging to high validation accuracies faster than with the classification data set. The learning rate was therefore slowed down to 0.000001, which produced a smoother curve for the loss and validation output. All models were trained up to 500 epochs, and achieved 100% Validation Accuracy.

5.4 RESULTS

	Precision	**Recall**	**F1**
VGG19	0.768701508	0.739130435	0.716299917
ResNet18	0.74781458	0.72463768	0.72458532
ResNet32	0.731562	0.71014493	0.71154978
ResNet50	0.81205534	0.7826087	0.78812315
ResNet101	0.76923741	0.73913043	0.73604777
ResNet152	0.92187716	0.89855072	0.90108407
DenseNet161	0.89452496	0.88405797	0.88204886

Table 8 – Identification Model Precision, Recall, and F1 Metrics

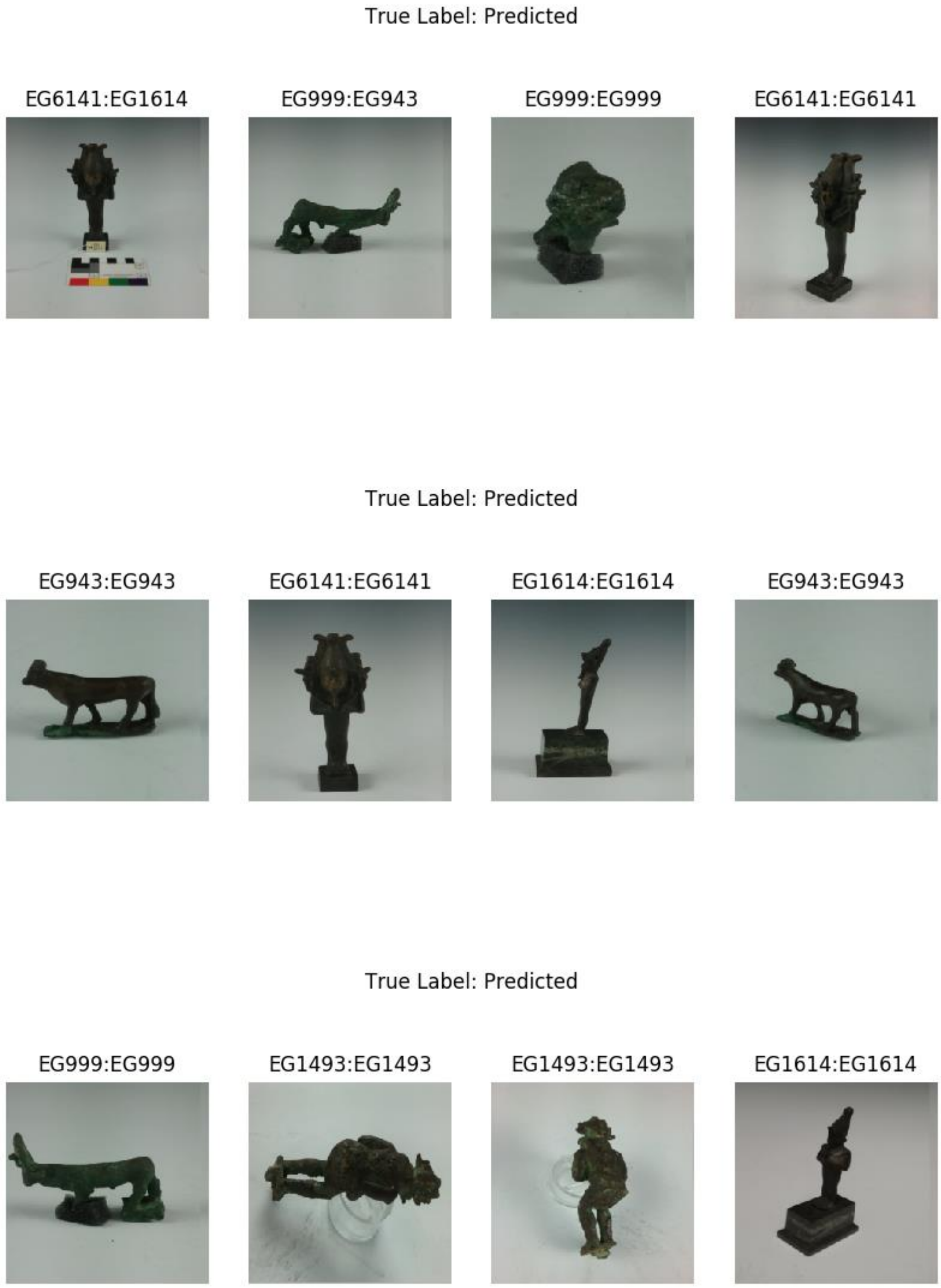

Figure 39 - Sample Output Predictions on HeDAP test set images, taken by Author.

5.4.1 VGGNet

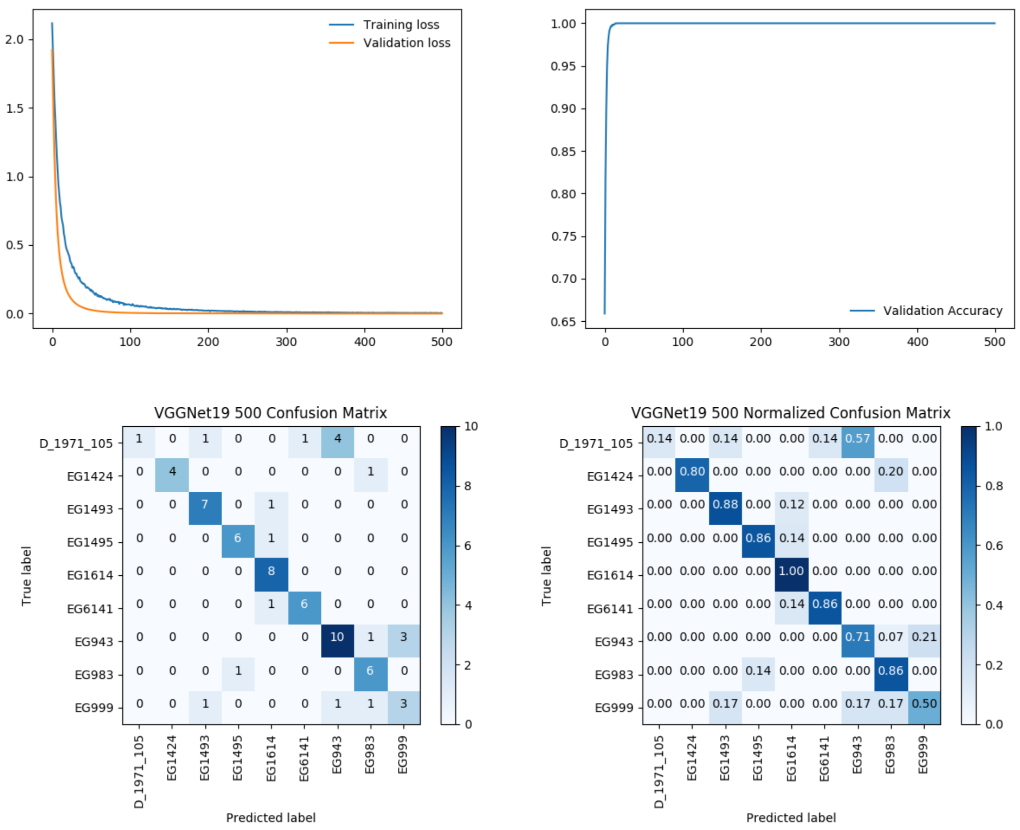

Figure 40 – VGGNet19 Single Object Identification Loss and Validation Graphs, Confusion Matrix Results (Non-normalized and Normalized)

5.4.2 ResNet Models

5.4.2.1 ResNet18

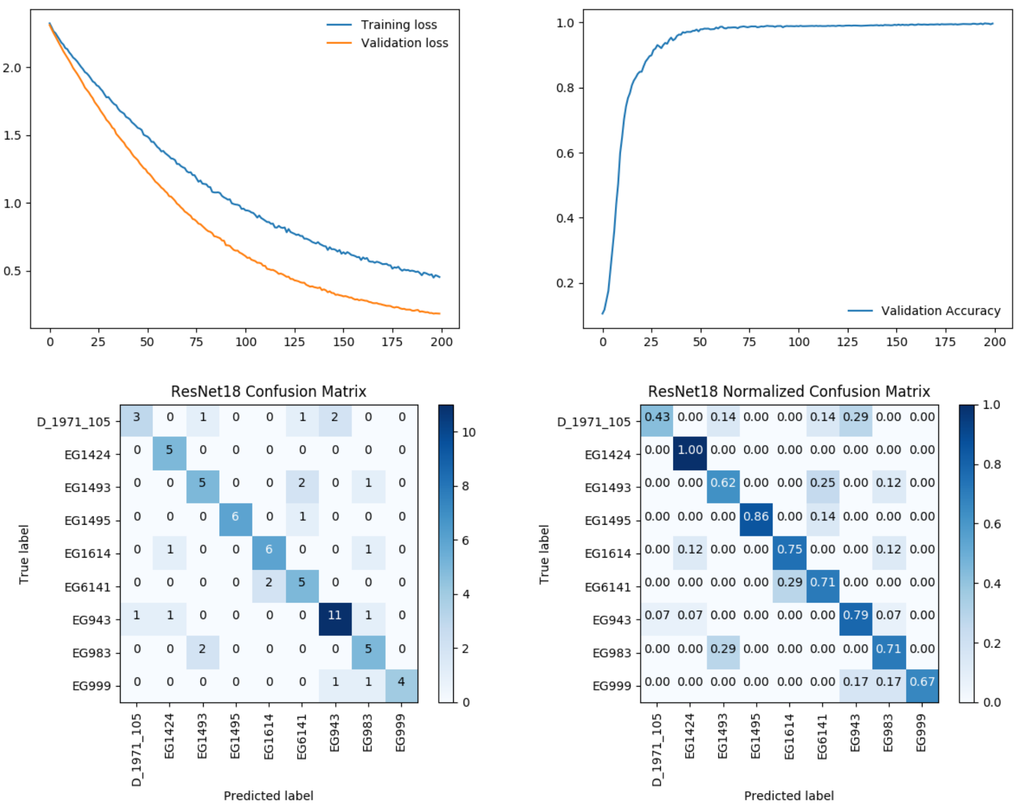

Figure 41 – ResNet18 Single Object Identification Loss and Validation Graphs, Confusion Matrix Results (Non-normalized and Normalized)

5.4.2.2 ResNet34

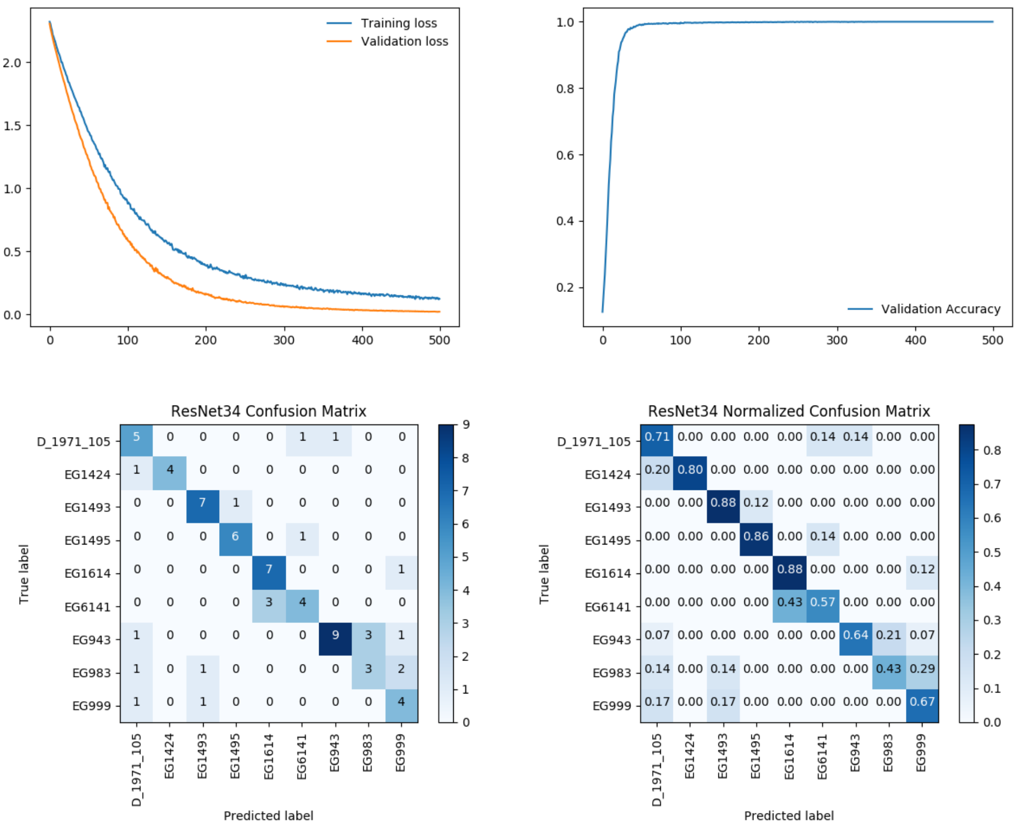

Figure 42 – ResNet34 Single Object Identification Loss and Validation Graphs, Confusion Matrix Results (Non-normalized and Normalized)

5.4.2.3 ResNet50

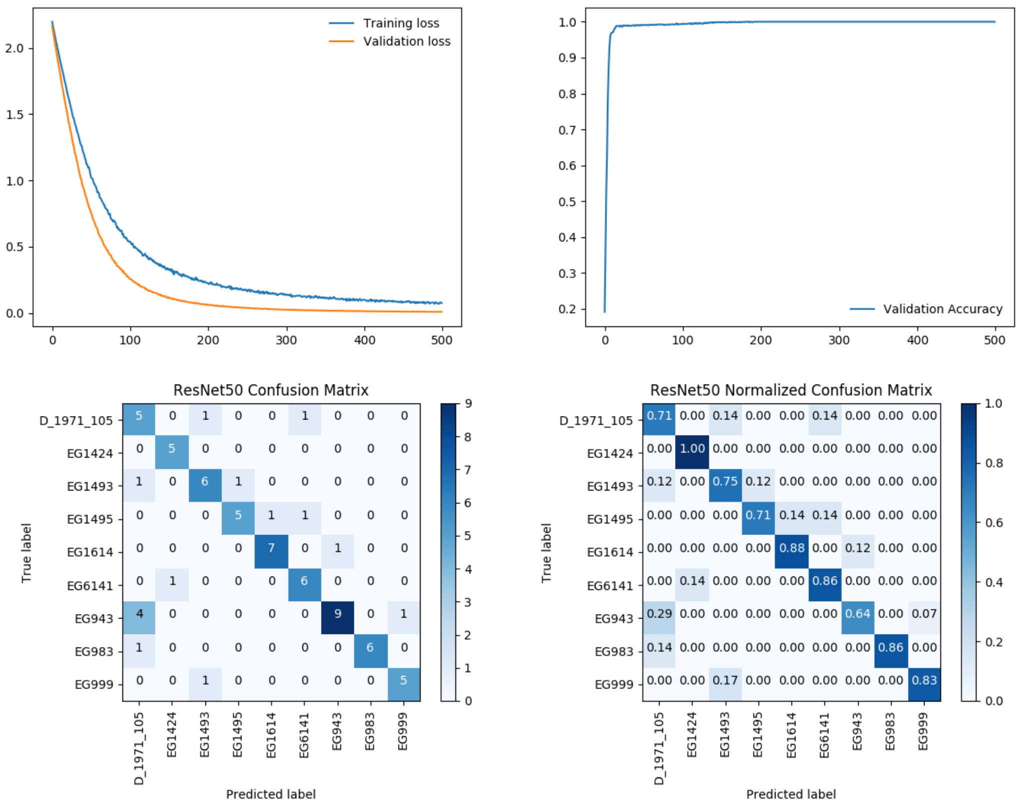

Figure 43 – ResNet50 Single Object Identification Loss and Validation Graphs, Confusion Matrix Results (Non-normalized and Normalized)

5.4.2.4 ResNet101

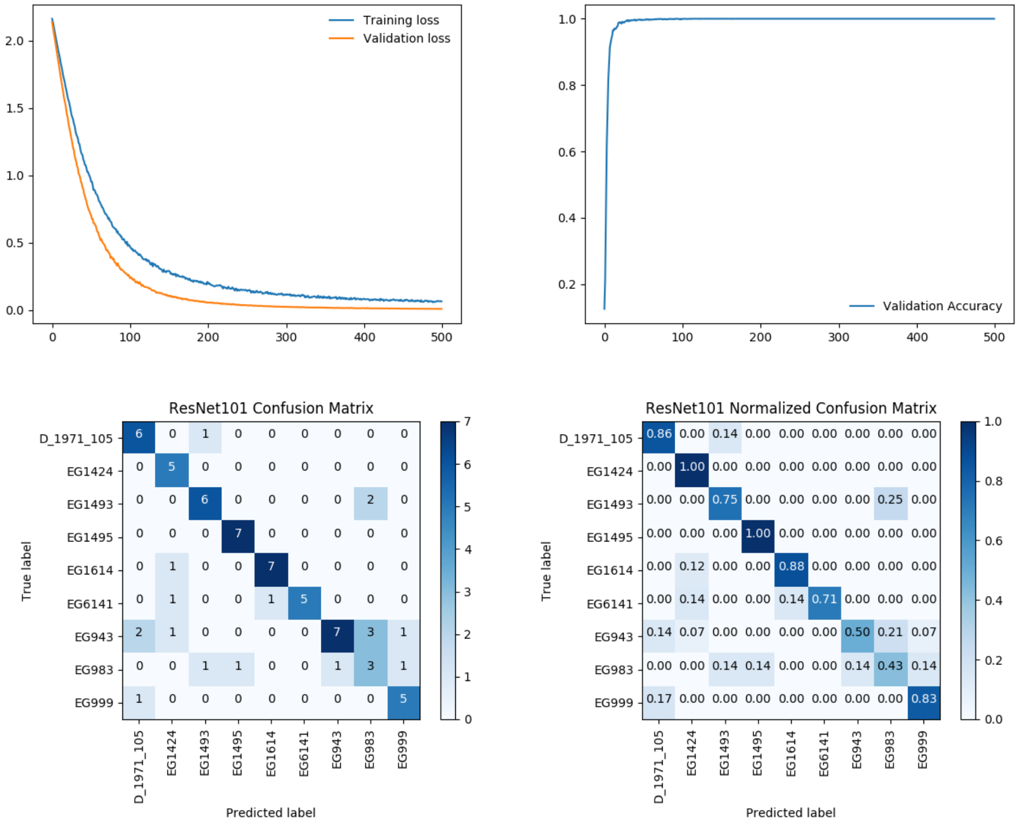

Figure 44 – ResNet101 Single Object Identification Loss and Validation Graphs, Confusion Matrix Results (Non-normalized and Normalized)

5.4.2.5 ResNet152

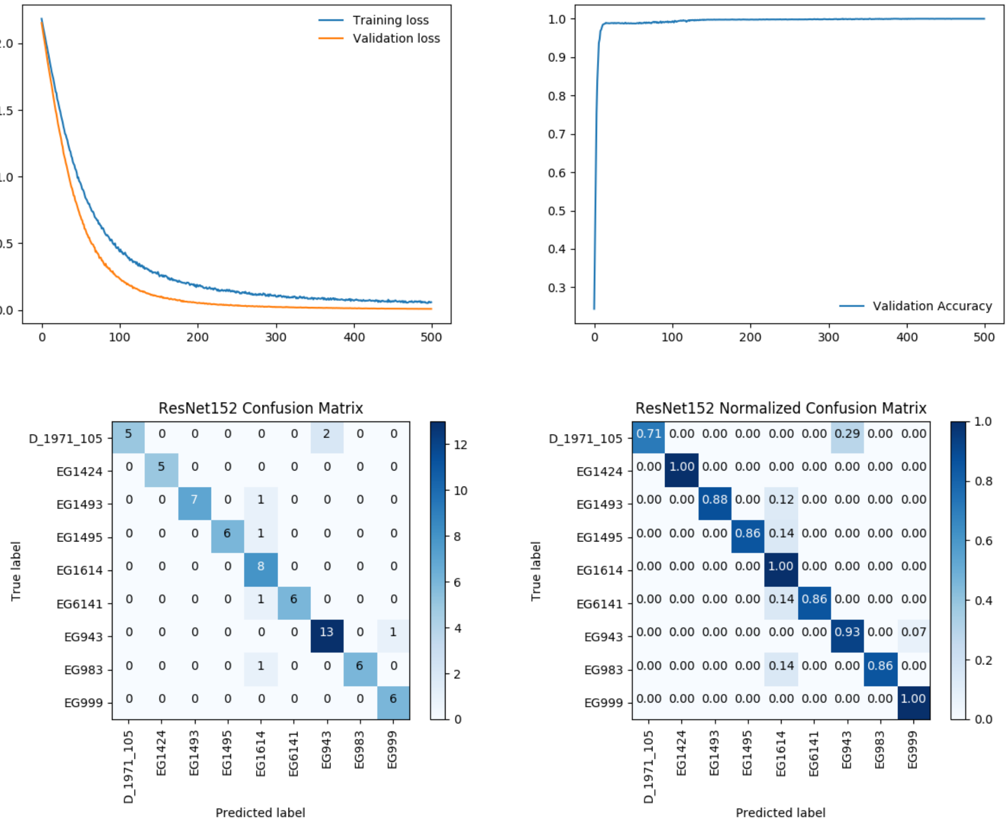

Figure 45 – ResNet152 Single Object Identification Loss and Validation Graphs, Confusion Matrix Results (Non-normalized and Normalized)

5.4.3 DenseNet

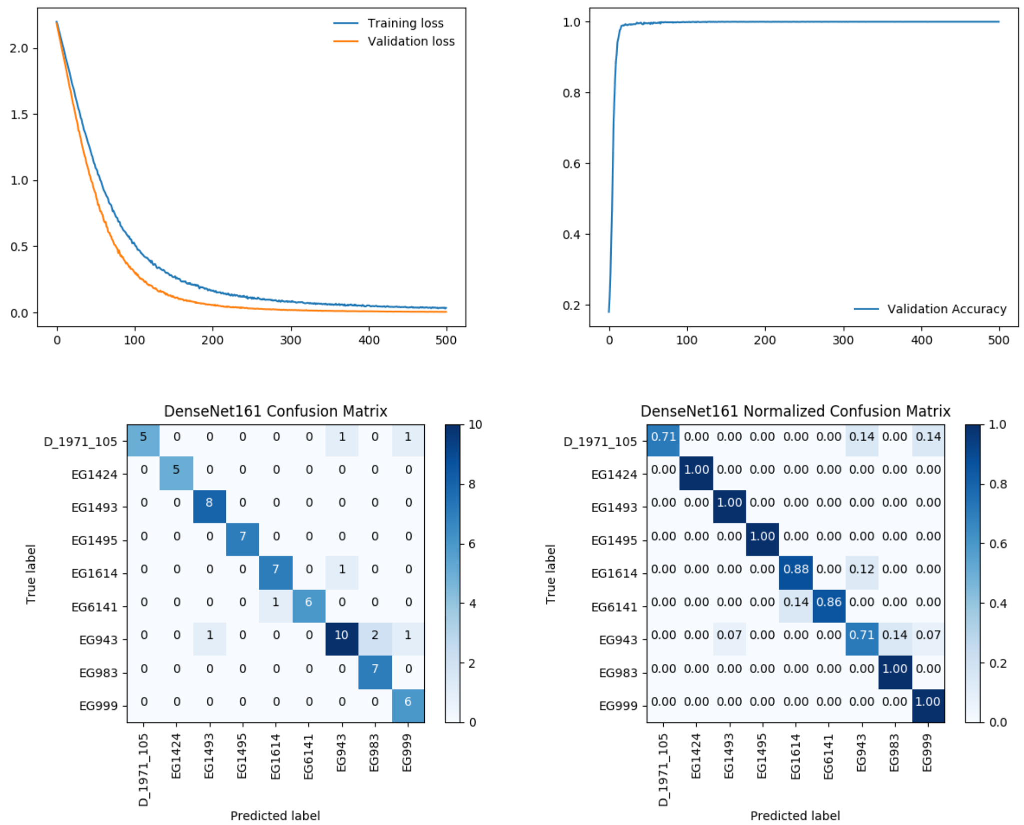

Figure 46 – DenseNet161 Single Object Identification Loss and Validation Graphs, Confusion Matrix Results (Non-normalized and Normalized)

5.5 Further Iteration: Combining Identification and Classification

Experimenting further, we sought to incorporate an element of classification into the training data. While Identification is the primary need in the HeDAP system description, a capability to also provide a suggested classification for an input would nevertheless be of some utility. Doing so would conceivably be similar to the ICOM Red Lists providing broad guidelines for typology match, even if a specific object match could not be found. Assessing the original Egyptian training data (see Figure 6), we chose to incorporate the ushabti image data. Not only does the ushabti data represent the highest concentration of a unique artefact type within our training classes, it is also a type of Egyptian artefact which is one of the most commonly found in museum collections, as well as sold through the antiquities trade (Stevenson 2015). Further, contained within the image data files provided by the Oriental Museum was the results of a student project to document all of the ushabti within their collection (see Figure 47). Consisting of 194 images, these photos contained front and back snapshots of the museum ushabti next to rulers and color markers. The background appears to be a mottled countertop, and the lighting and quality of photographs is highly variable across the images. Given that the HeDAP app also tends to produce lower-quality images than a DSLR camera, we felt that these images would represent a good test case for the added classifier.

Figure 47 - Ushabti Digital Images from Oriental Museum Records, source Durham University Oriental Museum.

Finally, we also chose to create a second test set of ushabti images scraped from the E-Bay auction site, inspired in part by the recent literature expressing concern about the effect of online market spaces on the illegal antiquities trade (as discussed in Section 2.2.4). Logging onto the site, we used the search term "ushabti" and collected images from the first page of auction results. Most included several front and back shots of ushabti, often being held by hand in order to provide context for size.

For our models we chose to focus on training ResNet152 and DenseNet161, due to their performance success from the Identification testing. We kept the same hyperparameter selection and backpropagation structure as described in Section 5.3.2. The training results can be seen below in Figure 48 and Figure 49. Even with the addition of the ushabti classifier set, the loss and accuracy graphs closely match the Identification training in the section above.

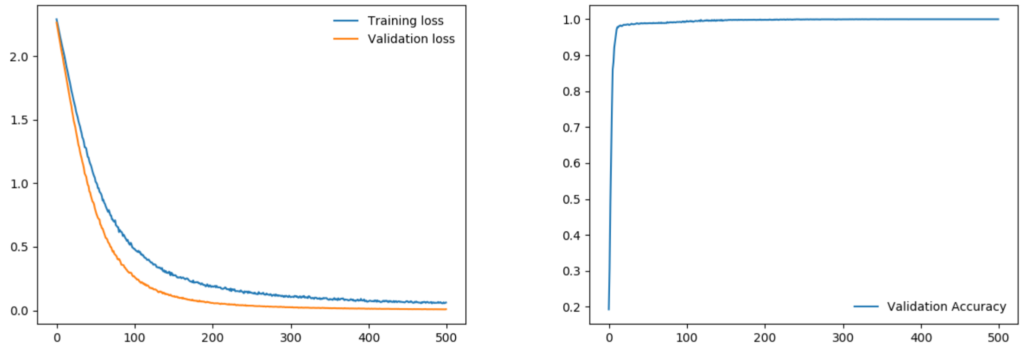

Figure 48 - ResNet152 Training Loss, Validation Loss, and Validation Accuracy

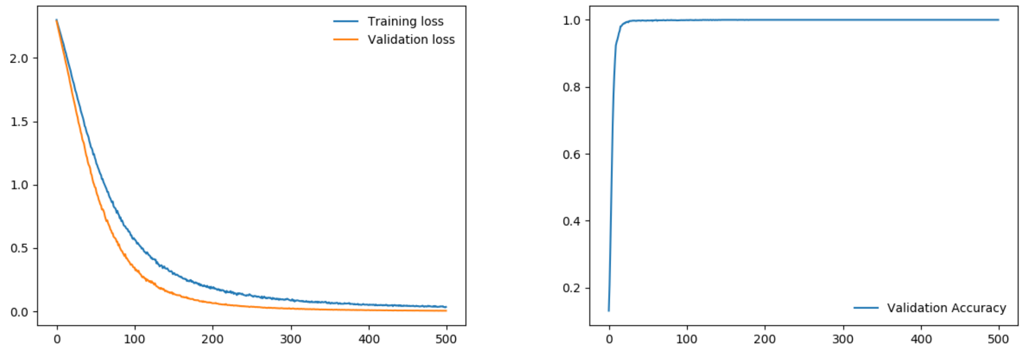

Figure 49 - DenseNet161 Training Loss, Validation Loss, and Validation Accuracy

5.5.1 ResNet152 Results

	Precision	Recall	F1
OM Ushabti Test Set	0.951300072973	0.946768060837	0.946166764311
eBay Ushabti Test Set	0.880582750583	0.855555555556	0.855706309468

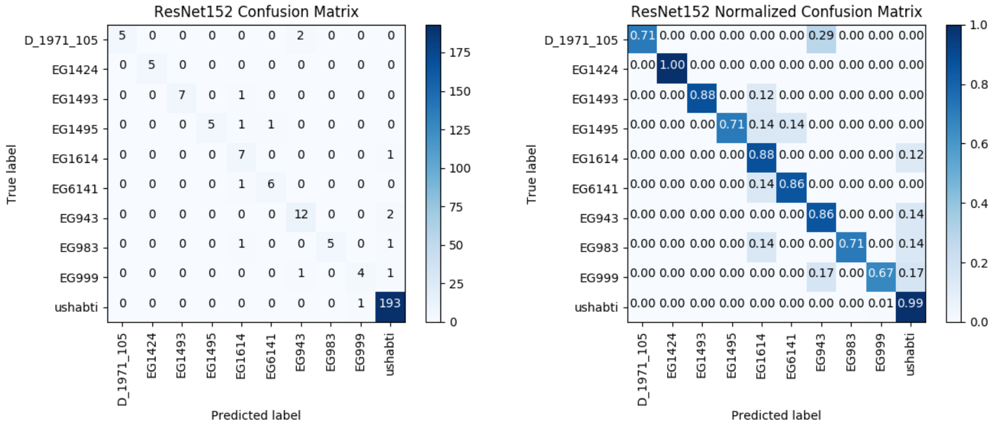

Figure 50 – ResNet152 Single Object Identification Combined with Ushabti Trained Model Confusion Matrix Results (Non-normalized and Normalized) for OM Ushabti Test Set

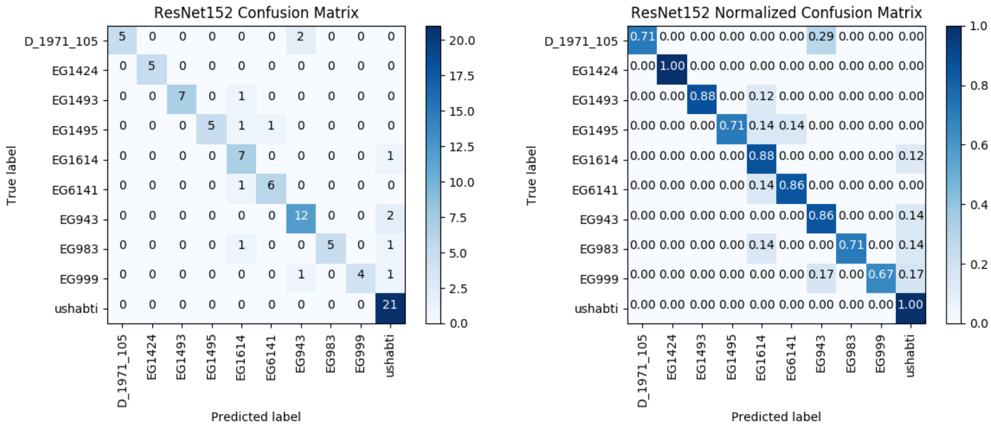

Figure 51 - ResNet152 Single Object Identification Combined with Ushabti Trained Model Confusion Matrix Results (Non-normalized and Normalized) for eBay Ushabti Test Set

5.5.2 DenseNet161 Results

	Precision	Recall	F1
OM Ushabti Test Set	0.985375442953	0.984790874525	0.984541728976
eBay Ushabti Test Set	0.948170594837	0.944444444444	0.944373800013

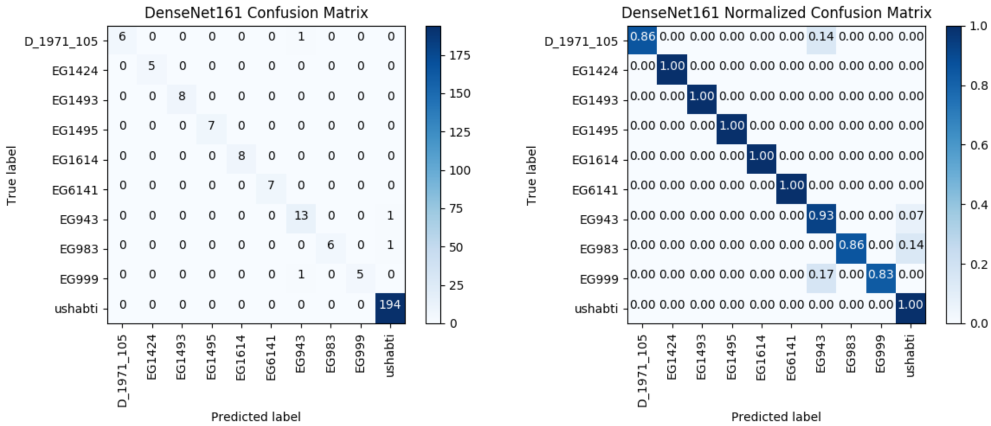

Figure 52 – DenseNet161 Single Object Identification Combined with Ushabti Trained Model Confusion Matrix Results (Non-normalized and Normalized) for OM Ushabti Test Set

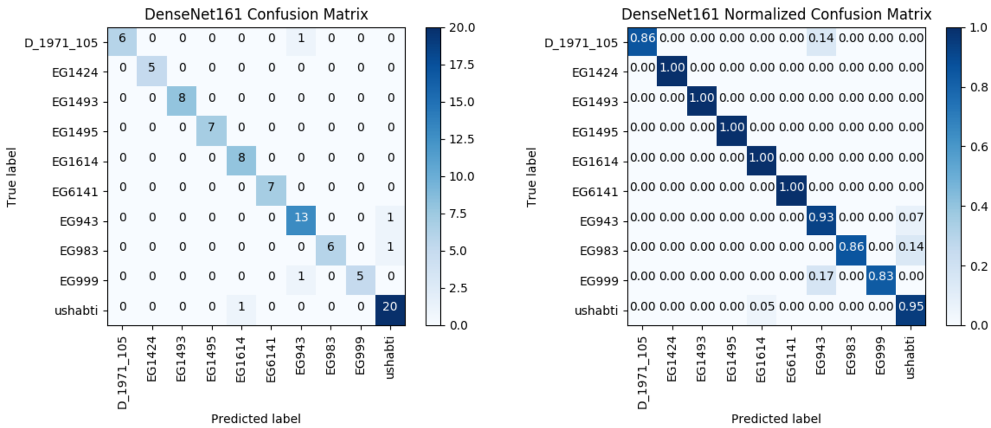

Figure 53 – DenseNet161 Single Object Identification Combined with Ushabti Trained Model Confusion Matrix Results (Non-normalized and Normalized) for eBay Ushabti Test Set

5.6 Discussion

In the case of Identification, we see that the deepest (i.e. greatest number of connected layers) models are generally performing better than the shallow models. As with the Classification testing, we found that ResNet50 outperforms ResNet101; however, it only achieves a 78.9% F1 score, compared to the ResNet152 and DenseNet161 scores of 90.1% and 88.2% accurate recognition, respectively. For the Validation and Accuracy Loss graphs, we do see in ResNet152 and DenseNet161 that the two losses seem closer to converging. Also, the Validation Accuracy graph (along with ResNet50) demonstrates a much sharper rise early-on in the training.

In regards to the addition of the ushabti Classification training, we found that DenseNet161 outperforms ResNet152. The metrics from the Oriental Museum student photography test set are skewed due to the imbalance between the museum data and the HeDAP Wellcome figurine test images (194 OM images vs 69 Wellcome images). However, even with this imbalance we still can still see that the DenseNet model is now outperforming the ResNet model, and more accurately assigning the individual bronze figurines into the appropriate category.

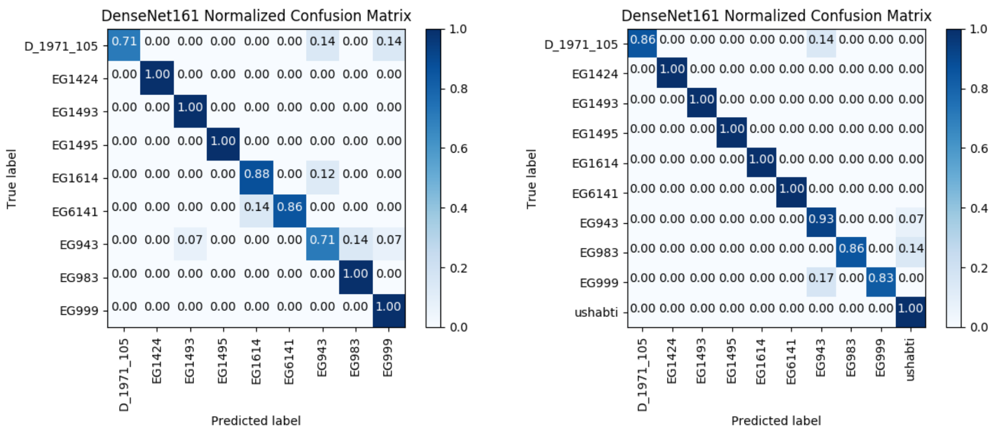

Figure 54 – DenseNet161 Single Object Identification Normalized Confusion Matrix compared to DenseNet161 Single Object Identification with Ushabti Training Normalized Confusion Matrix, the latter showing an increased capability of the model to accurately predict the individual figurines

Both models perform equally well classifying the ushabti images in all test sets. Suspecting that this might be too good a result, we created a new test set composed of 16 photos of dogs and cats using images from the Dogs vs Cats Kaggle competition dataset [8]. We labelled

[8] https://www.kaggle.com/c/dogs-vs-cats/overview

these images as ushabti, and ran it with our individual object test images through the DenseNet161 trained classifier. The final result was that all dog and cat photos were still predicted and labelled as ushabti (see Figure 55).

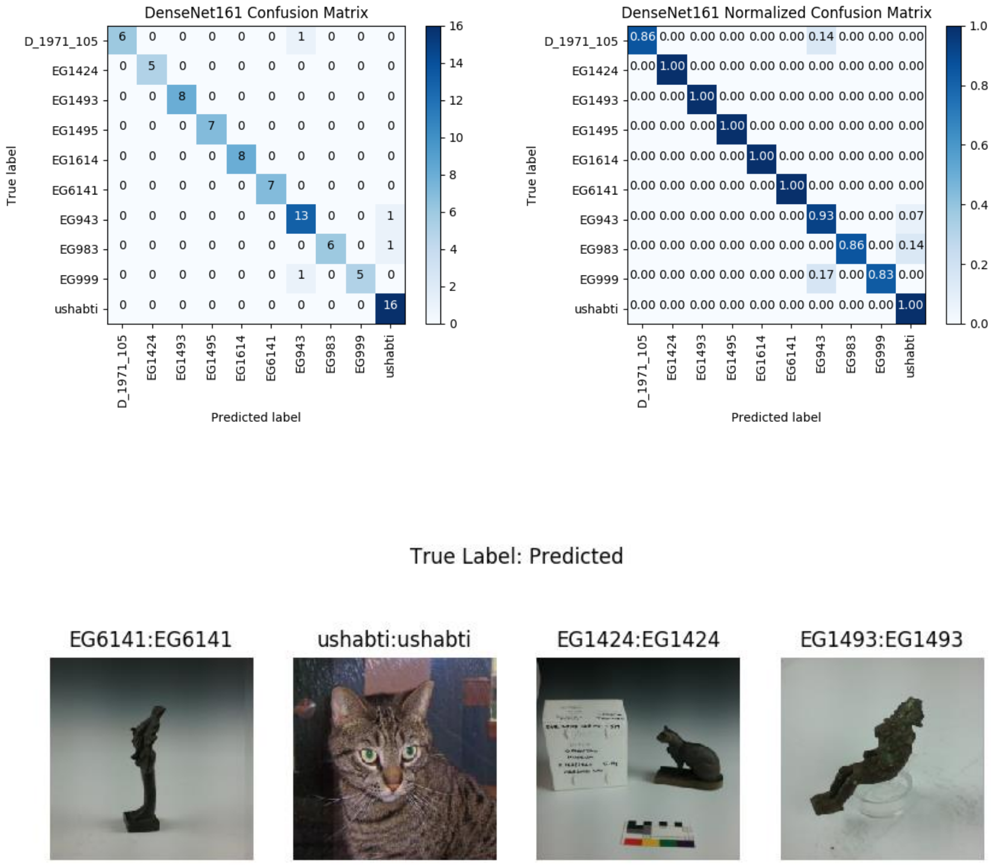

Figure 55 – DenseNet161 Single Object Identification Combined with Ushabti Trained Model Confusion Matrix Results (Normalized and Non-Normalized) and sample output for a test set of cat and dog test images labelled as ushabti

Thus while initially we considered the incorporation of the ushabti classifier a success, our new data results convinced us that the ushabti class was being used as a generic prediction for anything that didn’t match the internal filters the model had developed for the individual objects.

5.7 CONCLUSION

As with Object Classification, for Identification testing we continue to see a strong central line in the Confusion Matrix results. Also as with Classification, we found that uniformity of the training and test sets produces the high accuracy results. Our training set demonstrates that individual objects can function as a prediction class when digital image capture using burst mode photography is used to create image sets of 300 – 400 photos (roughly half the number used for the Six Class Typology training classes). Regarding the results stemming from adding in the ushabti class training, we found that the image data for the individual object identification is too uniform to be able to simply add in a new general style of object. Nevertheless, by adding the ushabti class image data we were able to improve the Identification results; we believe that this is due to the ushabti acting as a complicating factor to the training, thus forcing the machine learning optimization to find a more accurate set of filters to predict all of the training set. Thus even when constructing Identification data sets, consideration should be made for methods by which the models can be further challenged in order to develop a more robust set of prediction criteria.

One issue to consider is that the 100% Validation Accuracy might in fact be causing the models to Overfit (i.e. training a CNN model that matches the training data so closely that the model then fails to make correct predictions on new data). While the models perform well on the unseen image test sets taken in the Oriental Museum's photo lab (both for Classification and with the HeDAP Identification test set), their loss of predictive performance outside of the lab could be an indication that some other methodology should be used to control the training process. Regardless, the results indicate that multi-image photography of individual museum objects can be used to accurately identify a specific museum object. However, while potentially viable it requires a controlled photography setup, which may not be practical when deployed in an operational environment.

6 SUMMARY AND CONCLUSION

The automatic recognition of heritage material through computer vision and the use of convolutional neural networks has been presented in this work. The purpose has been framed in the context of the illicit antiquities market, with the discussion in Chapter 2 outlining the risk to heritage and the practical difficulties which could be addressed with an automated technical solution. By using a heritage institution's digital records, we created a new training data set of Egyptian and Asiatic figurines and vessels intended for the purposes of Classification. As described in Chapter 3, this was done in part due to not having enough source image data to focus solely on classification of Egyptian cultural material. The resulting training data set was tested using image sets of decreasing similarity, first using a holdout data set of identical background, lighting, image capture equipment, etc. and then progressing to test sets composed objects-in-the-wild and material from the same cultural grouping from other heritage institutions. In order to quantify the predictive results, we used common metrics and measures for judging image retrieval accuracy. We then implemented three architectures of pre-trained convnets in Pytorch (VGGNet, ResNet, and DenseNet) using Transfer Learning and the six-class Object Classification training set (documented in Chapter 4). Doing so resulted in very good (90+% Recall and Precision) results on the holdout test set using the largest models (ResNet152 and DenseNet161), as well as showing a significant capability for the more complex test sets (though with a notable performance loss for overly broad training classes). From this we were able to draw several conclusions related to our first research question (i.e. *how well does a curated, pre-existing multi-perspective image data set of a museum collection work for artefact typology classification?*). Ultimately we believe that the results show that a museum's digital records can be used for accurate predictions for broad classes and categories when based on similar and repeated features, when the digital records have been produced to a consistent documentation workflow. However, creation of coherent class groups are heavily dependent on a museum collection containing several copies of a particular type of objects, otherwise performance degradation could result from too broad categories.

As part of common machine learning development practices we then returned to the original problem and project goals, to extend the developed software further (detailed in

Chapter 5). Using the HeDAP project we considered the needs for a practical application of this work, specific to the needs of a system aimed at combatting the illicit trafficking of antiquities. Rather than Classification, the primary system goal was instead what we refer to as Identification (i.e. the creation of predictive classes consisting of a single specific artefact). Using burst mode photography and a set of small, previously un-photographed Egyptian figurines from the Wellcome Collection we created a new Identification training set. A test set was created using the HeDAP tablet application's digital recording capability, in the Oriental Museum's photo lab. The results were similar to the Classification predictive metrics, with ResNet152 and DenseNet161 returning the highest Precision and Recall results (ResNet152 Precision/Recall 92.2%/89.9%, DenseNet161 89.5%/88.4%). Regarding our second research question (*Can multi-image photography of individual museum objects be used to accurately identify a specific museum object?*), the results indicate that this method could be used for individual object identification, though the data collection workflow necessary for such is not as easily implementable for rapid heritage documentation. Iterating further, we sought to combine Object Identification with Object Classification, using the Egyptian ushabti subset of the anthropomorphic figurine class training data. Initially we found extremely high predictive results from test sets built from variations of ushabti data (one test set containing over 100 images of ushabti photographed in poor lighting conditions, the other containing ushabti images downloaded from the eBay auction site). However when a spoiler test set composed of cat and dog images marked as ushabti was run through the system, we discovered that the ushabti prediction was being used for any image which fell outside the individual Wellcome figurine objects. While we did see improvements to the accuracy of the Identification predictions, ultimately more time and resources would be necessary to determine how to combine Classification and Identification into a cohesive automated vision system.

Our primary contribution from this research has been to demonstrate the potential capabilities of CNN Transfer Learning when used in conjunction with an established heritage image archive, providing quantitative measurements for a variety of test scenarios. Seeking to relate this to a real-world context, we have created a set of practical guidelines for the digital recording practices associated with the HeDAP project, applying the lessons learned from our data preparation. Originally based on the ObjectID photo recommendations, these guidelines have been updated and refined over time as a result of the work done within the Oriental Museum's photo lab, as well as the adapting their good practices for image labelling and database recording. Feedback from the Training in Action field users and the system testing of the Heritage Documentation and Protection Android App also provided necessary improvements to the guidelines. These guidelines are included in Appendix II.

Based on the above results, our general assessment is that heritage image data has a great potential utility for data science research, but that the practicalities of data set creation makes implementing a fully reliable automated recognition system difficult to achieve. Existing Collection Management tools contain data fields which could be used to associate classification labels to digital images, a rich source of material for predictive training sets in Machine Learning. In practice, image data from a single, pre-existing heritage collection is unlikely to contain the necessary amount of duplicate object types required for Object Classification. Further, the organization of such data sets still requires a significant investment of time and experience and is not something that can be done automatically. We found that it was possible to create high predictive accuracies on individual object identification using modern tablets and smart phone recording capabilities, but that too requires time and experience. The software necessary for implementing a trained CNN model is broadly available through freely available tools and training online, but the current state of the technology still requires a strong background in programming and statistics to fine-tune a working system solution.

Overall, the use of Machine Learning with cultural heritage image data has great future potential. However, it will require wider consensus regarding documentation and storage practices, along with greater information sharing between custodian institutions to truly produce an effective and practical solution. As with other cultural heritage initiatives, success will ultimately depend on antiquities being made an accessible resource for the many rather than the few.

Appendix I. WELLCOME EGYPTIAN FIGURINES

	Accession Number	Date	Dimensions	Material
	DUROM.1971.105	Late Period	Width 74 mm Height 72 mm Depth 25 mm	Bronze
	EG943	Late Period	Width 11 mm Height 33 mm Depth 45 mm	Bronze
	EG999	Late Period	Width 24 mm Height 78 mm Depth 75 mm	Bronze

Table 9 - Egyptian Bull Figurines

	Accession Number	Date	Dimensions	Material
	EG983	Late Period	Width 17 mm Height 63 mm Depth 9 mm	Bronze
	EG1495	Late Period	Height 177 mm Width 52 mm Depth 45 mm	Bronze
	EG1614	Late Period	Width 45 mm Height 167 mm Depth 68 mm	Bronze; Modern marble base

	Accession Number	Date	Dimensions	Material
	EG6141	Late Period	Width 57 mm Height 167 mm Depth 35 mm	Bronze

Table 10 - Egyptian Osiris Figurines

	Accession Number	Date	Dimensions	Material
	EG1424	Ptolmaic Period	Height 137 mm Width 54 mm Depth 115 mm	Faience
	EG1493	Late Period	Height 106 mm Width 33 mm Depth 50 mm	Bronze

Table 11 - Egyptian Unique Figurines

Appendix II. Digital Photography Guidelines

General Recommendations

When taking pictures, it is important to keep in mind that a good image is not necessarily of the quality required for publication in an exhibition catalogue. The importance of these image records are to allow for the proper identification of the object. Aesthetics are less important than "readability" in this case.

Background

- If the object is made up of bright or dark colors: a neutral-colored background should be used.
- If the object is made of glass or is light in color: dark backgrounds should be used.
- The background should be made up of a smooth surface. This will avoid misreading the object's contours.

Lighting

- Direct light is unadvisable. It is preferable to create an ambient light with artificial lighting placed around the object.
- When using a single source of light, it should come from the upper right.
- The light should be neutral in color. Avoid yellow light-bulbs. It is preferable to favor "daylight" lamps.
- Crystal objects might prove the hardest to photograph. It is best to try different dark backgrounds and lights coming from one side or from the lower side.

Viewpoints

- Avoid shadows from labels and other elements encroaching on the space of the object. Rulers and color-charts are good for the first photograph, but if possible should be situated in such a way that they may be cropped out if necessary, not covering any part of the object.
- The full shape and all the contours of each object should be clearly visible, if possible.
- Two-dimensional objects, paintings, prints, coins, etc., should be photographed from a 90º angle, without distortions, and from the center of the object.
- Three-dimensional objects (including bas-reliefs objects which have become fragmented or excised from their original context), must be taken in several shots at different angles (forward, profile, three-quarters and top). Should the bottom of the object have specificities, it should also be photographed.
- Objects belonging to a set must be photographed both together and individually.
- Photograph whenever possible details of inscriptions, repairs, damage, such as cracks, or any other distinctive features that will help to differentiate the object from similar items.
- Objects must be centred and occupy as much of the image space as possible
- The longer part of the object must be placed horizontally or vertically, but never diagonally.

Things to avoid

- Shots of groups of objects, with the exception of sets of objects that belong together.
- Lighting effects that obscure the details of the object: reflections, shadows, light changes, etc. Lighting should be as neutral as possible.
- Avoid brightly coloured, textured or patterned backgrounds that can be distracting.

HeDAP Workflow Checklist

- ☐ Always begin by first making sure you are not adding additional photos to an existing entry. A new entry is created using the **TAKE PHOTO** button from the **Sites** screen.
- ☐ Check that the color or tone of the **Background** does not clash with or obscure the object. Background should be selected based on the material and color of the object. For large objects, this might require draping a cloth sheet behind the object.
- ☐ Determine what **Lighting** is required. This depends on the type of object and its physical structure; if it has pronounced grooves or carvings upon its surface, raking light is preferable to direct. However, try to avoid hard shadows thrown across the body of the object.
- ☐ Your first photograph should be a picture of the object with an identifying tag (an **Inventory or Accession Number** is best) along with a **Photographic Scale** and **Color Chart**. These will help when reviewing photos taken using the App. Remove these for the next set of photos (even when repeating the same view).
- ☐ Select **Viewpoints** to photograph the object. Try to fill up as much of the frame as possible. **At least four photographs** taken from all sides of an object are preferable (front, back, left, right). If photographing a flat object (such as a painting or cloth), front and back are fine. If possible, top and bottom photographs are also encouraged. **More photos is better than less**, as it gives a better chance of identification. Additional viewpoints can be selected based on the type of object, as well documenting **Distinguishing Features**.
- ☐ **Position** the tablet; ideally the object should move rather than the tablet (if possible). **Do not use the zoom function**; it only creates lower quality images. Instead, move the tablet itself closer to the object. After you have positioned the tablet to photograph the object, tap the object on the screen to **Focus** it.
- ☐ If you are not using a stand to hold the tablet in place, **Breathe Out** while taking the photograph of the object, and try not to move for at least a full second. This reduces the chance of blurring from motion.
- ☐ When done, if possible take time to **Review** the photos taken using the App in a darkened area to better view the screen without glare from surrounding lights. Exit out from the object entry details, to prepare for the next object.

BIBLIOGRAPHY

Anderson, David S. 2019. "Rampant Antiquities Theft Threatens Cultural Heritage Around The World." *Forbes*, July 31. Accessed November 11, 2019. https://www.forbes.com/sites/davidanderson/2019/07/31/rampant-antiquities-theft-threatens-cultural-heritage-around-the-world/.

Armstrong, Helen, interview by Matthew Roberts. 2018. *The Northumberland Collection* (February 26).

Arts Council England. 2014. "Designated Outstanding Collections." *Arts Council England.* Accessed 2019. https://www.artscouncil.org.uk/sites/default/files/download-file/Designated_Outstanding_collections.pdf.

Bar, Yaniv, Noga Levy, and Lior Wolf. 2014. "Classification of Artistic Styles Using Binarized Features Derived from a Deep Neural Network." *European Conference on Computer Vision Workshops.* Cham/Germany: Springer. 71-84.

Barclay, Craig, and Rachel Barclay. 2018. "'Filling the Gaps': Reconsidering the Contribution of Henry, Earl Percy, to the Alnwick Castle Collection of Egyptian Antiquities." *Cahiers Egypte Nilotique et Mediterraneenne* 19: 23-41.

Bator, Paul M. 1982. "An Essay on the International Trade in Art." *Stanford Law Review* 34: 277-384.

Bay, Herbert, Beat Fasel, and Luc Van Gool. 2006. "Interactive Museum Guide: Fast and Robust Recognition of Museum Objects." *Proceedings of the First International Workshop On Mobile Vision.* Capri, Italy: Association of Computing Machinery. 1-15.

Bengio, Yoshua, Pascal Lamblin, Dan Popovici, and Hugo Larochelle. 2007. "Greedy Layer-Wise Training of Deep Networks." *Advances in Neural Information Processing Systems* 153-160.

Benton, John F, Alan R Gillespie, and James M Soha. 1979. "Digital Image-Processing Applied to the Photography of Manuscripts." *Scriptorium* 33 (1): 40-55.

Birch, Samuel. 1880. *Catalogue of the Collection of Egyptian Antiquities at Alnwick Castle.* London: R. Clay Sons and Taylor printers.

Blighe, Michael, Sorin Sav, Hyowon Lee, and Noel E. O'connor. 2008. "Mo Músaem Fíorúil: A Web-Based Search and Information Service for Museum Visitors." *International Conference Image Analysis and Recognition.* Berlin, Heidelberg: Springer. 485-496.

Brindley, Lynne. 2002. "The Future of Libraries and Humanities Research: New Strategic Directions for the British Library." *Libraries & Culture* 26-36.

Brodie, Neil. 2012. "Organigram." *Trafficking Culture.* August 21. Accessed November 1, 2019. https://traffickingculture.org/encyclopedia/case-studies/organigram/.

Brodie, Neil. 2015. "The Internet Market in Antiquities." *Countering Illicit Traffic in Cultural Goods: The Global Challenge of Protecting the World's Heritage* 11-20.

Brodie, Neil. 2012. "Uncovering the Antiquities Market." In *The Oxford Handbook of Public Archaeology*, by Robin Skeates, Carol McDavid and John Carman, 230–52.

Brodie, Neil, and Jennifer Doole. 2001. "Illicit Antiquities." In *Trade in Illicit Antiquities: the Destruction of the World's Archaeological Heritage*, edited by Neil Brodie, Jennifer Doole and Colin Renfrew, 1-6. McDonald Institute for Archaeological Research.

Brodie, Neil, Jenny Doole, and Peter Watson. 2000. *Stealing History: The Illicit Trade in Cultural Material.* Cambridge: The McDonal Institute for Archaeological Research.

Burton, Virginia, Alice Bonin, Janice Lourie, and Toby Spiselman. 1970. "The Computer and Archaeology." *American Journal of Archaeology* 74 (3): 221–223.

Campbell, Peter B. 2013. "The Illicit Antiquities Trade as a Transnational Criminal Network: Characterizing and Anticipating Trafficking of Cultural Heritage." *International Journal of Cultural Property* 20 (2): 113-153.

Coggins, Clemency. 1969. "Illicit Traffic of Pre-Columbian Antiquities." *Art Journal* 29 (1): 94-114.

Collections Trust. 2017. "Spectrum." *Collections Trust.* September. Accessed 2019. https://collectionstrust.org.uk/spectrum/.

Cullen, John F, Jonathan J Hull, and Peter E. Hart. 1997. "Document Image Database Retrieval and Browsing Using Texture Analysis." *Proceedings of the Fourth International Conference on Document Analysis and Recognition.* Ulm, Germany: IEEE Computer Society. 718-721.

da Silva Torres, Ricardo, and Alexandre X. Falcao. 2006. "Content-Based Image Retrieval: Theory and Applications." *Revista de Informatica Teorica e Aplicada (RITA)* 13 (2): 161–185.

Dik, Joris, Milko den Leeuw, Wilko Verbakel, Ren´e Peschar, and Robert Schillemans. 2002. "The Digital Reconstruction of a Smalt Discoloured Painting by Hendrick Ter Brugghen." *Zeitschrift filer Kunsttechnologie und Konservierung* 16: 130-146.

Engel, Dulcie. 2017. "Henry Wellcome's Egyptian Legacy." *Y Ganolfan Eifftaidd / Egypt Centre.* September 20. Accessed June 01, 2019. http://www.egypt.swan.ac.uk/henry-wellcomes-egyptian-legacy-dulcie-engel/.

Föckler, Paul, Thomas Zeidler, Benjamin Brombach, Erich Bruns, and Oliver Bimber. 2005. "PhoneGuide: Museum Guidance Supported by On-Device Object Recognition on Mobile Phones." *Proceedings of the 4th international conference on Mobile and ubiquitous multimedia.* Christchurch, New Zealand: Association for Computing Machinery. 3-10.

Frossard, Davi. 2016. "VGG in TensorFlow." *University of Toronto Computer Science .* June 17. Accessed November 15, 2019. https://www.cs.toronto.edu/~frossard/post/vgg16/.

Gerges, Fawaz A. 2017. *ISIS: a History.* Princeton University Press.

Gerstenblith, Patty. 2013. "The Meaning of 1970 for the Acquisition of Archaeological Objects." *Journal of Field Archaeology* 38 (4): 364-373.

Giakoumis, Ioannis, and Ioannis Pitas. 1998. "Digital Restoration of Painting Cracks." *Proceedings of the 1998 Institute of Electrical and Electronics Engineers International Symposium on Circuits and Systems.* Orlando, USA: IEEE. 269-272.

Gonzalez, Rafael C., and Richard E. Woods. 2006. *Digital Image Processing (3rd Edition).* New Jersey (USA): Prentice-Hall, Inc.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. 2016. *Deep learning.* MIT Press.

Grigonis, Hillary. 2019. "How To Use Burst Mode for Perfectly-Timed Shots." *ExpertPhotography.com.* August. Accessed November 12, 2019. https://expertphotography.com/burst-mode-photos/.

Guerin, Orla. 2019. "Libya in Chaos as Endless War Rumbles On." *BBC News.* October 27. Accessed November 12, 2019. https://www.bbc.co.uk/news/world-africa-49971678.

Gulmohamad, Zana Khasraw. 2014. "The Rise and Fall of the Islamic State of Iraq and Al-Sham (Levant) ISIS." *Global Security Studies* 5 (2): 1-11.

Haladova, Zuzana. 2010. "Segmentation and Classification of Fine Art Paintings." *14th Central European Seminar on Computer Graphics.* Budmerice, Slovakia: Impressum. 59-65.

Hamber, Anthony. 2006. "Digitisation in the commercial art world." In *Digital heritage: Applying digital imaging to cultural heritage*, by Lindsay MacDonald, 115-147. Oxford: Elsevier.

Hare, Jonathon S., Paul H. Lewis, Peter GB Enser, and Christine J. Sandom. 2006. "Mind the Gap: Another Look at the Problem of the Semantic Gap in Image Retrieval." *Multimedia Content Analysis, Management, and Retrieval 2006.* International Society for Optics and Photonics. 607309.

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. "Deep Residual Learning for Image Recognition." *Proceedings of the Institute of Electrical and Electronics Engineers Conference on Computer Vision and Pattern Recognition.* Institute of Electrical and Electronics Engineers. 770-778.

—. 2015. "Delving Deep Into Rectifiers: Surpassing Human-Level Performance on Imagenet Classification." *Proceedings of the Institute of Electrical and Electronics Engineers International Conference on Computer Vision.* Institute of Electrical and Electronics Engineers. 1026-1034.

Hinton, Geoffrey E., Simon Osindero, and Yee-Whye Teh. 2006. "A Fast Learning Algorithm for Deep Belief Nets." *Neural Computation* 18 (7): 1527-1554.

Howarth, Frank. 2012. "World Museum Leadership: ICOM or AAM?" *Museum Management and Curatorship* 27 (2): 87-92.

Huang, Gao, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q. Weinberger. 2017. "Densely Connected Convolutional Networks." *Proceedings of the Institute of Electrical and Electronics Engineers Conference on Computer Vision and Pattern Recognition.* Institute of Electrical and Electronics Engineers. 4700-4708.

ICOM Department of Programmes. 1954. "Convention for the Protection of Cultural Property in the Event of Armed Conflict with Regulations for the Execution of the Convention 1954." *The ICOM International Observatory on Illicit Traffic in Cultural Goods.* May 14. https://www.obs-traffic.museum/sites/default/files/ressources/files/UNESCO_convention_1954.pdf.

ICOM. 2012. "Statement of principles of museum documentation." *ICOM International Committee for Museums and Collections of Archaeology and History.* June. Accessed 2019. http://network.icom.museum/fileadmin/user_upload/minisites/cidoc/DocStandards/Statement6v2EN.pdf.

International Color Consortium. 1996. "ICC Profile Format Specification." *International Color Consortium* 1: 1998-09.

International Council of Museums. 2019. "Red Lists." *International Council of Museums.* Accessed September 2, 2019. https://icom.museum/en/activities/heritage-protection/red-lists/.

Jordan, Michael I, and Tom M Mitchell. 2015. "Machine learning: Trends, perspectives, and prospects." *Science* (American Association for the Advancement of Science) 349: 255-260.

Kaiser, Timothy. 1991. "The Antiquities Market." *Journal of Field Archaeology* 87-103.

Kersel, Morag M. 2007. "Transcending Borders: Objects on the Move." *Archaeologies* 3 (2): 81-98.

Keskar, Nitish Shirish, and Richard Socher. 2017. "Improving Generalization Performance by Switching from Adam to SGD." *arXiv preprint arXiv:1712.07628.*

Kiernan, Kevin S. 1991. "Digital Image Processing and the Beowulf Manuscript." *Literary and Linguistic Computing* 6 (1): 20-27.

Kim, Jongpil, and Vladimir Pavlovic. 2016. "Discovering Characteristic Landmarks on Ancient Coins using Convolutional Networks." *Journal of Electronic Imaging* 26 (1): 011018.

Kingma, Diederik P., and Jimmy Ba. 2014. "Adam: A Method for Stochastic Optimization." *arXiv:1412.6980v9.*

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. 2012. "Imagenet Classification with Deep Convolutional Neural Networks." *Advances in Neural Information Processing Systems* 1097-1105.

Larson, Frances. 2009. *An Infinity of Things: How Sir Henry Wellcome Collected the World.* Oxford University Press.

LeCun, Yann, Léon Bottou, Yoshua Bengio, and Patrick Haffner. 1998. "Gradient-Based Learning Applied to Document Recognition." *Proceedings of the IEEE.* 2278-2324.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. 2015. "Deep Learning." *Nature* 521 (7553): 436.

Liu, Ying, Dengsheng Zhang, Guojun Lu, and Wei-Ying Ma. 2007. "A Survey of Content-Based Image Retrieval with High-Level Semantics." *Pattern Recognition* 40 (1): 262-282.

Mackenzie, Simon, and Tess Davis. 2014. "Temple Looting in Cambodia: Anatomy of a Statue Trafficking Network." *British Journal of Criminology* 54 (5): 722-740.

Martinez, Kirk. 1991. "High Resolution Digital Imaging of Paintings: the Vasari Project." *Microcomputers for Information Management* 8 (4): 277–283.

Martinez, Kirk, John Cupitt, David Saunders, and Ruven Pillay. 2002. "Ten Years of Art Imaging Research." *Proceedings of the Institute of Electrical and Electronics Engineers* 90 (1): 28–41.

Massy, Laurence. 2008. "The Antiquity Art Market: Between Legality and Illegality." *International Journal of Social Economics* 35 (10): 729-738.

Mensink, Thomas, and Jan Van Gemert. 2014. "The Rijksmuseum Challenge: Museum-Centered Visual Recognition." *Proceedings of International Conference on Multimedia Retrieval.* Glasgow/United Kingdom: Association of Computing Machinery. 451.

Meyer, David A. 1993. "The 1954 Hague Cultural Property Convention and its Emergence into Customary International Law." *Boston University International Law Journal* 11: 349-389.

Mugnai, Niccolò, Julia Nikolaus, David Mattingly, and Susan Walker. 2017. "Libyan Antiquities at Risk: Protecting Portable Cultural Heritage." *Libyan Studies* 48: 11-21.

Nafziger, James A.R. 1972. "Regulation by the International Council of Museums: an Example of the Role of Non-Governmental Organizations in the Transnational Legal Process." *The Denver Journal of International Law and Policy* 2 (2): 231-254.

Pan, Sinno Jialin, and Qiang Yang. 2009. "A Survey on Transfer Learning." *Institute of Electrical and Electronics Engineers Transactions on Knowledge and Data Engineering* 10 (22): 1345-1359.

Papert, Seymour A. 1966. "The Summer Vision Project." *MIT Computer Science and Artificial Intelligence Lab.* July 7. Accessed September 23, 2019. http://people.csail.mit.edu/brooks/idocs/AIM-100.pdf.

Paszke, Adam, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. 2017. "Automatic Differentiation in Pytorch." *31st Annual Conference on Neural Information Processing Systems.* Long Beach, USA. Accessed September 10, 2019.

Penn Museum. 1980. "The Pennsylvania Declaration." *Expedition Magazine* 22 (3): 3. Accessed August 5, 2019. https://www.penn.museum/sites/expedition/the-pennsylvania-declaration/.

Percy, Earl Henry Algernon George Percy. 1901. *Highlands of Asiatic Turkey.* London: Edward Arnold.

Pezzati, Alessandro. 2010. "The Pennsylvania Declaration." *Expedition*, 4. Accessed 2019. https://www.penn.museum/documents/publications/expedition/PDFs/52-2/from%20the%20archives.pdf.

Polk, Kenneth. 2014. "The Global Trade in Illicit Antiquities: Some New Directions?" In *Heritage Crime*, edited by Louise Grove and Suzie Thomas, 206-223. London: Palgrave Macmillan.

Prescott, Andrew. 1997. "The Electronic Beowulf and Digital Restoration." *Literary and Linguistic Computing* 12 (3): 185-196.

Prott, Lyndel V. 1996. "UNESCO and UNIDROIT: a Partnership Against Trafficking in Cultural Objects." *Uniform Law Review-Revue de Droit Uniforme* 1 (1): 59-71.

Rajalingham, Rishi, Elias B. Issa, Pouya Bashivan, Kohitij Kar, Kailyn Schmidt, and James J. DiCarlo. 2018. "Large-Scale, High-Resolution Comparison of the Core Visual Object Recognition Behavior of Humans, Monkeys, and State-of-the-Art Deep Artificial Neural Networks." *Journal of Neuroscience* 38 (33): 7255-7269.

Rawat, Waseem, and Zenghui Wang. 2017. "Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review." *Neural computation* 29 (9): 2352-2449.

Redies, Christoph, Anselm Brachmann, and Johan Wagemans. 2017. "High Entropy of Edge Orientations Characterizes Visual Artworks from Diverse Cultural Backgrounds." *Vision Research* 133: 130-144.

Reppel, Erik. 2017. "Visualizing parts of Convolutional Neural Networks using Keras and Cats." *Hackernoon.* January 22. Accessed November 10, 2019. https://hackernoon.com/visualizing-parts-of-convolutional-neural-networks-using-keras-and-cats-5cc01b214e59.

Republic of Cyprus Customs and Excise Department. 2019. "Announcements." *Ministry of Finance (Cyprus).* Accessed September 20, 2019. https://www.mof.gov.cy/mof/customs/customs.nsf/All/F59A7A3A9E451708C22579DC002813F1?OpenDocument&print.

Ruder, Sebastian. 2016. "An Overview of Gradient Descent Optimization Algorithms." *arXiv preprint arXiv:1609.04747.*

Ruf, Boris, Effrosyni Kokiopoulou, and Marcin Detyniecki. 2008. "Mobile Museum Guide Based On Fast SIFT Recognition." *International Workshop on Adaptive Multimedia Retrieval.* Berlin, Germany: Springer. 170-183.

Ruffle, John. 1998. "The Journeys of Lord Prudhoe and Major Orlando Felix in Egypt, Nubia and the Levant 1826-1829." In *Travellers in Egypt*, by Paul Starkey and Janet Starkey, 75-84. I.B. Tauris & Co Ltd.

Ruiz, Pablo. 2018. "Understanding and Visualizing DenseNets." *Towards Data Science.* October 10. Accessed November 15, 2019. https://towardsdatascience.com/understanding-and-visualizing-densenets-7f688092391a.

Russell, Georgina. 1987. "The Wellcome Historical Medical Museum's Dispersal of Non-Medical Material, 1936-1983." *Museums Journal* 86: 21-45.

Ryan, Nick S. 1988. "A Bibliography of Computer Applications and Quantitative Methods in Archaeology." *Computer and Quantitative Methods in Archaeology.* Oxford: British Archaeology Reports. 1-27.

Schlag, Imanol, and Ognjen Arandjelovic. 2017. "Ancient Roman Coin Recognition in the Wild using Deep Learning Based Recognition of Artistically Depicted Face Profiles." *International Conference on Computer Vision Workshop.* Venice/Italy: Institute of Electrical and Electronics Engineers. 2898–2906.

Shao, Ling, Fan Zhu, and Xuelong Li. 2015. "Transfer Learning for Visual Categorization: A Survey." *Institute of Electrical and Electronics Engineers Transactions on Neural Networks and Learning Systems* 26 (5): 1019-1034.

Shearer, Colin. 2000. "The CRISP-DM Model: the New Blueprint for Data Mining." *Journal of Data Warehousing* 5 (4): 13-22.

Simonyan, Karen, and Andrew Zisserman. 2014. "Very Deep Convolutional Networks for Large-Scale Image Recognition." *arXiv preprint arXiv:1409.1556.*

Smeulders, Arnold WM, Marcel Worring, Simone Santini, Amarnath Gupta, and Ramesh Jain. 2000. "Content-Based Image Retrieval at the End of the Early Years." *Institute of Electrical and Electronics Engineers Transactions on Pattern Analysis and Machine Intelligence* 22 (12): 1349-1380.

Sokolova, Marina, Nathalie Japkowicz, and Stan Szpakowicz. 2006. "Beyond Accuracy, F-score and ROC: a Family of Discriminant Measures for Performance Evaluation." *Australasian Joint Conference on Artificial Intelligence* (Springer) 1015-1021.

Srivastava, Nitish, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. "Dropout: a Simple Way to Prevent Neural Networks from Overfitting." *The Journal of Machine Learning Research* 15 (1): 1929-1958.

Stanco, Filippo, Alfredo Restrepo, and Giovanni Ramponi. 2011. "Virtual Restoration of Antique Books and Photographs." *Digital Imaging for Cultural Heritage Preservation: Analysis, Restoration, and Reconstruction of Ancient Artworks* 263-298.

Stevenson, Alice. 2015. "Egyptian Archaeology and the Museum." *Oxford Handbooks Online.* February 3. Accessed September 23, 2019. https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199935413.001.0001/oxfordhb-9780199935413-e-25.

Tan, Wei Ren, Chee Seng Chan, Hernán E. Aguirre, and Kiyoshi Tanaka. 2016. "Ceci N'est Pas Une Pipe: A Deep Convolutional Network for Fine-Art Paintings Classification." *International Conference on Image Processing.* Phoenix/USA: Institute of Electrical and Electronics Engineers. 3703-3707.

Tansey, Elizabeth M. 2002. "Medicines and Men: Burroughs, Wellcome & Co, and the British Drug Industry Before the Second World War." *Journal of the Royal Society of Medicine* 95 (8): 411-416.

Taylor, Matthew E., and Peter Stone. 2009. "Transfer Learning for Reinforcement Learning Domains: A Survey." *Journal of Machine Learning Research* 10: 1633-1685.

Thacker, Thomas W. 1949. "The Alnwick Collection of Egyptian Antiquities." Oriental Museum Archive, April 30.

Thornes, Robin, Peter G. Dorrell, Henry Lie, and Peter Lie. 2000. *Introduction to Object ID: Guidelines for Making Records that Describe Art, Antiques, and Antiquities.* Getty Publications.

Tijhuis, Antonius Johannes Gerhardus. 2009. "Trafficking in Cultural Artifacts." In *The Oxford Handbook of Crime and Public Policy*, edited by Michael Tonry, 455-474. Oxford University Press, Inc.

Trustees of the British Museum. 1835. "The Prudhoe Lions." *The British Museum.* Accessed November 19, 2019. https://research.britishmuseum.org/research/collection_online/collection_object_details.aspx?assetId=775539001&objectId=117626&partId=1.

UNESCO. 1970. "Convention on the Means of Prohibiting and Preventing the Illicit Import, Export and Transfer of Ownership of Cultural Property 1970." *UNESCO.ORG.* November 14. http://portal.unesco.org/en/ev.php-URL_ID=13039&URL_DO=DO_TOPIC&URL_SECTION=201.html.

UNIDROIT. 2014. *Overview - UNIDROIT Convention On Stolen or Illegally Exported Cultural Objects.* June. https://www.unidroit.org/overviecp/english.

United Nations Security Council. 2015. "Resolution 2199." *United Nations Documents.* February 12. Accessed September 2, 2019. https://www.undocs.org/S/RES/2199%20(2015).

—. 2015. "Resolution 2253." *United Nations Documents.* December 17. Accessed September 2, 2019. https://www.undocs.org/S/RES/2253%20(2015).

—. 2017. "Resolution 2347." *United Nations Documents .* March 24. Accessed September 2, 2019. https://www.undocs.org/S/RES/2347%20(2017).

Valentino-DeVries, Jennifer. 2011. "May the Best Algorithm Win…." *Wall Street Journal*, March 16.

Vikhrov, Natalie. 2018. "Stolen Heritage: How Russia is Destroying Crimean Archaeological Sites." *Hromadske International.* January 4. Accessed September 20, 2019. https://en.hromadske.ua/posts/stolen-heritage-how-russia-is-destroying-crimean-archaeological-sites.

Warkworth, Henry Algernon George. 1898. *Notes from a Diary in Asiatic Turkey, by Lord Warkworth.* London: Edward Arnold.

Watson, Peter. 1997. *Sotheby's: the Inside Story.* London: Bloomsbury.

Watson, Peter, and Cecilia Todeschini. 2007. *The Medici Conspiracy.* PublicAffairs.

Weiss, Karl, Taghi M. Khoshgoftaar, and DingDing Wang. 2016. "A Survey of Transfer Learning." *Journal of Big Data* 3 (9): 1-40.

Wilcock, John D. 1973. "A General Survey of Computer Applications in Archaeology." *Science and Archaeology* (9): 17-21.

Wilcock, John D. 1999. "Getting the Best Fit? 25 Years of Statistical Techniques in Archaeology." *Computer Applications and Quantitative Methods in Archaeology 1997.* Birmingham: British Archaeology Reports. 35-52.

Wilkinson, John Gardner. 1837. *The Manners and Customs of the Ancient Egyptians.* London: John Murray.

Wilson, Ashia C., Rebecca Roelofs, Mitchell Stern, Nati Srebro, and Benjamin Recht. 2017. "The Marginal Value of Adaptive Gradient Methods in Machine Learning." *Advances in Neural Information Processing Systems* 4148-4158.

World Customs Organization. 2019. "Membership." *World Customs Organization Organisation Mondiale des Douanes.* Accessed September 2, 2019. http://www.wcoomd.org/en/about-us/wco-members/membership.aspx.

—. 2005. "Model export certificate for cultural objects: explanatory notes." *UNESDOC Digital Library.* Accessed September 20, 2019. https://unesdoc.unesco.org/ark:/48223/pf0000139620.

—. 2019. "Publications." *World Customs Organization Organisation Mondiale des Douanes.* Accessed September 2, 2019. http://www.wcoomd.org/en/topics/enforcement-and-compliance/resources/publications.aspx.

—. 2012. "WCO Illicit Trade Report 2012." *World Customs Organization Organisation Mondiale des Douanes.* Accessed September 20, 2019. http://www.wcoomd.org/en/topics/enforcement-and-compliance/resources/publications.aspx.

—. 2015. "WCO Illicit Trade Report 2015." *World Customs Organization Organisation Mondiale des Douanes.* Accessed September 20, 2019. http://www.wcoomd.org/en/topics/enforcement-and-compliance/resources/publications.aspx.

—. 2016. "WCO Illicit Trade Report 2016." *World Customs Organization Organisation Mondiale des Douanes.* Accessed September 20, 2019. http://www.wcoomd.org/en/topics/enforcement-and-compliance/resources/publications.aspx.

—. 2017. "WCO Illicit Trade Report 2017." *World Customs Organization Organisation Mondiale des Douanes.* Accessed September 20, 2019. http://www.wcoomd.org/en/topics/enforcement-and-compliance/resources/publications.aspx.